IMFUFA tekst

- I, OM OG MED MATEMATIK OG FYSIK

On the teaching and learning of probability and statistics in the perspective of Critical Mathematics Education

Raimundo José Elicer Coopman
PhD thesis

nr. 513 – 2020

Roskilde University,
Department of Science and Environment, IMFUFA
P.O. Box 260, DK - 4000 Roskilde
E-mail: imfufa@ruc.dk

On the teaching and learning of probability and statistics in the perspective of Critical Mathematics Education

Raimundo José Elicer Coopman
PhD thesis

IMFUFA tekst nr. 513/2020 — 269 sider — ISSN: 0106-6242

This is a shortened abstract. Full abstract available in main text.

In this thesis, I define and address research questions about the justification, possibility and implementation problem fields of mathematics education, anchored in Chilean upper secondary school context; to explore the coherence along a *problématique* of teaching and learning probability and statistics through the lens of Critical Mathematics Education.

In a comparative study, I first investigate the role of decision making in Chilean and Danish probability and statistics upper secondary school curricula. The main finding is that, though it is acknowledged that scenarios of uncertainty and statistical information have societal relevance, the connection to probability and statistics as school subjects is unclear beyond the personal psychological scope.

Concerning the design, I delineate three design principles, namely exemplarity, inquiry approach and pragmatism. They account for the quality of the chosen examples, the structure of the learning environments, and practical considerations of the teaching context. As a result of four classroom experiences, some general tensions and re-definitions are made.

I draw on Skovsmose's notion of reflective knowing to analyse the ability of the learning environments to provoke critical reflections on the formatting power of mathematics in the classroom.

As for the domain-specific character of students' reflective knowing, five overarching probabilistic and statistical constructs play a significant role in those reflections and are approachable in upper secondary school: the *conflicting meanings* of probability, the assumption of *independence*, the asymmetry and meanings of *conditional probabilities*, the use of *distributions* to make sense of phenomena, and statistical *inference* as an inductive tool.

Raimundo José Elicer Coopman

On the teaching and learning of probability and statistics in the perspective of Critical Mathematics Education

Dissertation for the degree *Philosophiæ Doctor (PhD)*
This thesis has been submitted to the Doctoral School of Science and Environment

Supervisor: Morten Blomhøj

Roskilde University
Department of Science and Environment
November 2020

Abstract

Critical Mathematics Education (CME) can be characterised in terms of preoccupations related to issues of power in mathematics and mathematics education. In particular, it acknowledges the formatting power of mathematics, by which applications of mathematics give shape and functioning to aspects of society. In particular, probability and statistics play a particular role in the mathematisation of uncertainty, risks and decisions, as well as making sense of variation and data. Decision problems have been attached to probability and statistics throughout their historical development, as the scope of decisions has widened from games of chance to scientific, economic, political and social matters of the world.

In this thesis, I define and address research questions about the justification, possibility and implementation problem fields of mathematics education, anchored in Chilean upper secondary school context; to explore the coherence along a *problématique* of teaching and learning probability and statistics through the lens of CME.

In a comparative study, I first investigate the role of decision making in Chilean and Danish probability and statistics upper secondary school curricula. Employing a version of Fairclough's Critical Discourse Analysis, I focus on two discursive practices, namely the positioning of social actors and legitimation strategies. The analysed texts reflect a notion of students as future – not current– rational and data-based decision-makers, enabled by what is learned in school. The inclusion of probability and statistics in curricula is legitimised by narratives of rationality and appeals to authority, both political and academic. The main finding is that, though it is acknowledged that scenarios of uncertainty and statistical information have societal relevance, the connection to probability and statistics as school subjects is unclear beyond the personal psychological scope.

The rest of the thesis concerns the challenges of designing and implementing a teaching-learning practice that aims to be coherent with the critical perspective on probability and statistics education. The overall strategy of the thesis is an exploratory multiple-case study, where cases are defined as four classroom experiences taking place in different upper secondary school classrooms, in collaboration with three teachers in Chile.

"Guilty or innocent?", based on a real case of wrongful conviction, allows students to work with the notions of data and probabilistic evidence, and to discuss the problematic use of probability in matters of the law. In "How are PSU scores computed?", students explore the calculation of higher education admission tests, illustrating their background and foreground through percentiles, social statistics in a fabricated normality, and political arithmetic in graphs. In "Should we install a thermoelectric power plant?", students are confronted to a decision utilising probability calculations, exploring the complexities involved in scenarios of risk. Finally, in "How many people attended the students' march?", a crowd size estimation task, students

experience the construction and sampling of data, the use of statistical estimation, and the role of context knowledge. These four learning environments inform research questions regarding considerations for design, students' critical reflections and the specificity of probability and statistics in those reflections.

Concerning the design, I delineate three design principles, namely exemplarity, inquiry approach and pragmatism, accounting for the quality of the chosen examples, the structure of the learning environments, and practical considerations of the teaching context. As a result of the classroom experiences, some general tensions and re-definitions are made. A case can be *exemplary* from at least three perspectives; subjective, instrumental and critical. These can enter in conflict depending on what it is prioritised; closeness to students personal experiences, probabilistic and statistical disciplinary ideas, or how these ideas shape the world. More importantly, exemplarity must be experienced and so becomes updated as a quality of the teaching. The *inquiry approach* has students work as mathematicians and scientists do, with a guiding inquiry put forward. The authenticity of the tasks can be problematic since real-world problems can disengage students from questions that do not relate to their lives and eventually overwhelm the curriculum. Adaptations may be appropriate but can miss the purpose of reflecting upon the real out-of-school applications of mathematics. The experiences show that students' independent work must be balanced with support and crystallisation of mathematical ideas that become necessary to advance the inquiry. *Pragmatism* should be seen not as a strict set of restrictions but as a flexible set of practical considerations that presupposes constant and trustful negotiation with the teacher. Above all, the tensions help to re-define the design principles as design dimensions.

In terms of students' activity, I draw on Skovsmose's notion of reflective knowing to analyse the ability of the learning environments to provoke critical reflections on the formatting power of mathematics in the classroom. Based on the analyses and the teacher's input, I layout four main issues upon which students *did* our *could* reflect.

First, the assumption of independence can validate the collection of data and shape our perceptions of risk. Second, probabilistic calculations and constructs that rely on frequentist and *a priori* interpretations play a role of conveying objective degrees of belief. Probability affects our claims of knowledge. Third, statistical investigations need to operationalise categories and variables to make sense of variation. When applied to the social world, these categories can produce discourses of political divide and exclusion of certain kinds of people. Fourth, probability and statistics not only shape perceptions but inform automated and deliberate action; their usage has consequences. Moreover, other topics related to the preoccupations of CME emerge as well, which may or may not be only related to probability and statistics. These are the notion of foreground, the fact that data are constructed, and the attention to decision-makers and stakeholders.

As for the domain-specific character of students' reflective knowing, I make use of different frameworks for analysis on each case: a framework specific to the problem of probability and the law, a proto-framework for statistical graphs developed with the students, a framework for probability thinking and literacy in the context of risk, and a framework for critical statistical literacy. As a result, five overarching probabilistic and statistical constructs play a significant role and are approachable in upper secondary school. Probability can oversimplify the perception of a variety of *conflicting meanings* into one single metaphor. The puzzling theoretical character of *independence* is a necessary assumption to make the uncertainty of complex systems mathematically graspable. *Conditional probabilities* cluster different interpretations of non-independence into one asymmetric construct, seldom possible to reverse, and conveys statements of probabilistic evidence and association. *Distributions* give sense to phenomena by providing shape to order, variation and uncertainty, using theoretical probabilistic models or arrangements of empirical data. Statistical *inference* provides formal procedures to grasp the uncertainty involved in the inductive process of using sample data to make generalisations about a population.

Resumen en español

La Educación Matemática Crítica (EMC) puede ser caracterizada en términos de preocupaciones relacionadas con temas de poder en la matemática y en la educación matemática. En particular, reconoce el poder formateador de la matemática, a través del cual aplicaciones de las matemáticas dan forma y funcionamiento a aspectos de la sociedad. La probabilidad y la estadística juegan un rol particular en la matematización de la incertidumbre, riesgo y decisiones, así como en darle sentido a la variabilidad y los datos. Problemas de decisión han estado adheridos a la probabilidad y estadística a lo largo de su desarrollo histórico, mientras el campo de aplicación se ha ampliado desde los juegos de azar hasta asuntos científicos, económicos, políticos y sociales del mundo.

En esta tesis defino y abordo preguntas de investigación relacionadas con los campos de problemas de justificación, posibilidad e implementación de la educación matemática, anclados en el contexto de tercer y cuarto año de enseñanza media chilena; para explorar la coherencia a lo largo de la problemática de la enseñanza y aprendizaje de probabilidad y estadística a través del lente de la EMC.

A través de un estudio comparativo, primero investigo el rol de la toma de decisiones en el currículo secundario superior de probabilidad y estadística en Chile y Dinamarca. Empleando una versión de Análisis Crítico del Discurso de Fairclough, me enfoco en dos prácticas discursivas, a saber, el posicionamiento de actores sociales y estrategias de legitimización. Los textos analizados reflejan una noción de estudiantes como futuros – no actuales – tomadores de decisiones racionales y basadas en datos, habilitados por lo que se aprende en la escuela. La inclusión de probabilidad y estadística en los currículos se legitima a través de narrativas de racionalidad y apelación a la autoridad, tanto política como académica. El principal hallazgo es que, aunque se reconoce que escenarios de incertidumbres y con información estadística tienen relevancia social, la conexión con probabilidad y estadística como materias escolares no es clara, más allá del foco personal y psicológico.

El resto de la tesis trata de los desafíos en diseñar e implementar una práctica de enseñanza-aprendizaje que apunta a ser coherente con la perspectiva crítica sobre la educación en probabilidad y estadística. La estrategia general de la tesis es un estudio de casos múltiples, donde los casos los definen cuatro experiencias de aula. Éstas tuvieron lugar en diferentes aulas de enseñanza media, en colaboración con tres profesores en Chile.

"¿Culpable o inocente?", basada en un caso real de condena errónea, permite a los estudiantes trabajar con nociones de datos y evidencia probabilística, y discutir el uso problemático de probabilidad en asuntos de la ley. En "¿Cómo se calculan los puntajes PSU?", los estudiantes exploran el cálculo de pruebas de admisión a la educación superior, ilustrando sus antecedentes (*background*) y *foregorunds*, la estadística social en una normalidad fabricada, y la aritmética

política en gráficos. En "¿Deberíamos instalar una planta termoeléctrica?", los estudiantes son confrontados a una decisión utilizando cálculos de probabilidades, explorando las complejidades involucradas en escenarios de riesgo. Finalmente, en "¿Cuántas personas asistieron a la marcha estudiantil?", una actividad de estimación de multitudes, los estudiantes experimentan la construcción y muestreo de datos, el uso de estimación estadística, y el rol del conocimiento contextual. Estos cuatro ambientes de aprendizaje informan preguntas de investigación respecto a consideraciones para el diseño, reflexiones críticas de los estudiantes y la especificidad de la probabilidad y la estadística en dichas reflexiones.

Con respecto al diseño, delineo tres principios, a saber, ejemplaridad, enfoque indagatorio y pragmatismo, dando cuenta de la calidad de los ejemplos elegidos, la estructura de los ambientes de aprendizaje, y consideraciones prácticas del contexto de enseñanza. Como resultado de las experiencias de aula, se muestran algunas tensiones generales y redefiniciones.

Un caso puede ser *ejemplar* desde al menos tres perspectivas; subjetiva, instrumental y crítica. Éstas pueden entrar en conflicto dependiendo de qué se prioriza; la cercanía las experiencias personales de los estudiantes, las ideas disciplinarias de probabilidad y estadística, o cómo estas ideas moldean el mundo. Más importantemente, la ejemplaridad debe ser vivenciada y así se actualiza como una cualidad de la enseñanza. El *enfoque indagatorio* sitúa a los estudiantes trabajando como lo hacen matemáticos y científicos, mediante una indagación guía establecida desde un principio. La autenticidad de las actividades puede ser problemática, ya que problemas del mundo real pueden desconectar a los estudiantes de preguntas que no se relacionan con sus vidas y que eventualmente superan al currículo. Adaptaciones pueden ser apropiadas, pero pueden perder el propósito de reflexionar sobre las aplicaciones de las matemáticas fuera de la escuela. Las experiencias muestran que el trabajo independiente de los estudiantes se puede equilibrar con apoyo y la cristalización de ideas matemáticas que se hacen necesarias para avanzar en la indagación. El *pragmatismo* no debiese verse como un conjunto de restricciones estrictas sino como un conjunto de consideraciones prácticas que presuponen una negociación constante y confiable con el profesor. Sobre todo, las tensiones ayudan a redefinir los principios de diseño como dimensiones de diseño.

En cuanto a la actividad de los estudiantes, me valgo de la noción de saber reflexivo de Skovsmose para analizar la capacidad de los ambientes de aprendizaje para provocar reflexiones críticas sobre el poder formateador de la matemática en el aula. Basado en el análisis y en las impresiones de los profesores, trazo cuatro asuntos principales sobre los que los estudiantes *reflexionaron* o *podrían* reflexionar. Primero, el supuesto de independencia puede validar la recolección de datos que moldean nuestras percepciones de riesgo. Segundo, cálculos y constructos de probabilidad que se valen de las interpretaciones frecuentistas y *a priori* juegan un rol de transmitir grados objetivos de creencia. La probabilidad afecta nuestras afirmaciones de conocimiento. Tercero, las investigaciones estadísticas necesitan operacionalizar categorías y variables para dar sentido a la variabilidad. Cuando se aplican al mundo social, estas categorías pueden producir discursos de división política y de exclusión de ciertos tipos de personas. Cuarto, la probabilidad y la estadística no solo moldean percepciones, sino que informan acciones automáticas y deliberadas; su uso tiene consecuencias. Más aún, otros tópicos relacionados con las preocupaciones de la EMC emergen también, que pueden o no estar relacionados con probabilidad y estadística. Éstos son la noción de *foreground*, el hecho de que los datos se construyen, y la atención a los tomadores de decisiones y partes interesadas.

Respecto al carácter de dominio específico del saber reflexivo de los estudiantes, hago uso de diferentes marcos para el análisis en cada caso: un marco específico para el problema de probabilidad y la ley, un proto-marco para gráficas estadísticas desarrollado con los estudiantes, un marco para el pensamiento y alfabetización estadísticos en contextos de riesgo, y un marco para la alfabetización estadística crítica. Como resultado, cinco constructos probabilísticos y estadísticos generales juegan un rol significativo y son accesibles durante la enseñanza media. La probabilidad puede sobre simplificar la percepción de una diversidad de *significados en conflicto* en una metáfora única. El carácter teórico de la *independencia* es un supuesto necesario para hacer manejable la incerteza en sistemas complejos. Las *probabilidades condicionales* agrupan diferentes interpretaciones de la no-independencia en un solo constructo asimétrico, raramente posible de invertir, y transmite afirmaciones de evidencia probabilística y de asociación. Las distribuciones dan sentido a fenómenos a través de darle forma al orden, variación e incertezas, usando modelos probabilísticos teóricos o arreglos de datos empíricos. La *inferencia* estadística provee procedimientos formales para asir la incertidumbre involucrada en el proceso inductivo de usar datos muestrales para hacer generalizaciones acerca de una población.

Acknowledgements

First of all, I would like to give special thanks to all students that participated in this investigation, willing to go by this stranger's ideas. Their work, engagement and discussions made this project what it is. The empirical part of the project was possible because of the generous collaboration of the three teachers involved, which I refer to as Alejandra, Bernardo and Carla. I appreciate their trust in me getting into your classrooms and try something different, for better or worse. Any of the ideas of critically learning probability and statistics expressed in this thesis *are* and *could be* possible because of the noble profession you have chosen. All of your energy and commitment made me miss the classrooms as well.

To a large extent, the discussion in Chapter 3 is a follow-up to an essay assignment from the PhD course *Historical and epistemological perspectives on mathematics education* at Stockholm University, Sweden. My appreciation and admiration go to professors Paola Valero and Iben Maj Christiansen for leading such profound yet friendly conversations. Special thanks to Kristina Palm Kaplan for her feedback and suggestions in that assignment.

A draft version of the design principles developed in Section 4.2 has been submitted as a paper co-authored by Morten Blomhøj, and accepted for presentation at the 14th International Congress on Mathematics Education (ICME-14). In particular, the examination of the exemplary principle in subsection 4.2.1 is based on a group project for the Certificate of University Teaching and Learning (CUTL) at Roskilde University, developed together with Mette Wichmand, Mette Blok and Rikke Haller Baggesen. Without them and our work in that project, I could not have dug into such a philosophical thought process about exemplarity, and let alone base it upon original texts in German and Danish.

A PhD is both an academic endeavour and an emotional journey. In that regard, I thank my family for their unconditional and patient love at a distance; my parents Francisco and Cecilia, and my siblings Francisco, Isabel, Maximiliano and Benjamín. Along the same line, thanks to my close friends in Chile and Denmark for your support. You know who you are.

I want to show my appreciation to the Department of Science and Environment at Roskilde University for being a friendly and welcoming environment for academic discussion. Listing all of the people that positively impacted my PhD life in the Department would deplete this page. However, I carry special consideration to my students, PhD fellows, lunch buddies, and musical accomplices. Special thanks to whom became my close friends: Lorenzo Costigliola, Corinna Schiano di Cola, Athina Andrea and Bartłomiej Kołaczkowski.

A special regard to Uffe Thomas Jankvist from *Danmarks institut for Pædagogik og Uddannelse* - DPU at Aarhus University and the group of PhD students he leads: Rikke Maagaard Gregersen, Ingi Heinesen Højsted, Cecilie Carlsen Bach, Mathilde Kjær Pedersen, Morten Elkær Hansen, Maria Kirstine Østergaard, Marianne Thomsen, Birgitte Henriksen, Dandan Sun and Christelle Plummer. Thank you all for your feedback, perspective, and sense of companionship and belonging.

Muchas gracias to my comrades Melissa Andrade-Molina, Alex Montecinos, Vanessa Franco Neto and Gustavo Bruno. They have been a real inspiration to join the conversation about what is critical in mathematics education in Latin America. Special thanks to Eduardo Carrasco for pushing me into this world and Mario Sánchez Aguilar for making it possible.

Above all, *tusind tak* to my supervisor Morten Blomhøj for believing in my personal adventure, often more than I did and factually before I could make sense of it. His constant advice, feedback, support, humour and wise guidance were necessary conditions for making this journey possible.

This work was funded by Chile's National Agency for Research and Development (ANID) / Scholarship Programme / DOCTORADO BECAS CHILE / 2016 - 72170530.

Raimundo José Elicer Coopman
Roskilde, Denmark
November 2020

Research outputs

This dissertation is formatted as a self-contained monograph. The following is an account of the written research outputs developed during the course of the PhD studies. Some of these papers are cited throughout the thesis as part of the reference list. The article Elicer (2020) is partially transcribed in Section 3.3, with permission of the editor.

Published

Elicer, R. (2020). Meanings of decision-making in probability and statistics: Comparing Chilean and Danish upper secondary school curricula. *Nordic Studies in Mathematics Education*, *25*(1), 25–44.

Elicer, R. (2019). The role of decision-making in the legitimation of probability and statistics in Chilean upper secondary school curriculum. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11, February 5 - 10, 2019)* (pp. 930–937). Utrecht, the Netherlands: Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Elicer, R., & Neto, V. (2018). Statistics as a tool for decision-making: Two countries and one pattern. In Gómez, D. M. (Ed.), *Proceedings of the First PME Regional Conference: South America*, p. 150. Rancagua, Chile: PME.

Franco Neto, V., Elicer, R., & Bruno, G. (2018). Mathematics education research from and in Latin America. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education (Vol. 1)* (pp. 203–204). Umeå, Sweden: PME.

Elicer, R. (2018). Statistical messages in Chile's presidential campaign as a concern for mathematics education. Paper presented at *JustEd2018 International Research Conference: Promoting Social Justice through Education*, (p. 76). Helsinki, Finland: Nordic Centre of Excellence - Justice through Education in the Nordic Countries.

Elicer, R., & Carrasco, E. (2017). Conditional probability as a decision-making tool: A didactic sequence. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME10, February 1 – 5, 2017)* (pp. 748–755). Dublin, Ireland: DCU Institute of Education and ERME.

Unpublished

Elicer, R. (2020). *A crowd size estimation task in the context of protests in Chile*. Paper accepted for presentation.

Elicer, R., & Blomhøj, M. (2019). *The critical stance in statistics didactical design*. Paper accepted for presentation.

Elicer, R. & Blomhøj, M. (2017, May-June). *Designing for students' decision-making and critical reflections: A tax paying game*. Paper presented at the Eighth Nordic Conference on Mathematics Education, Stockholm, Sweden.

Acronyms

APT	A Priori Theory of Probability
ASA	American Statistical Association
CDA	Critical Discourse Analysis
CME	Critical Mathematics Education
FQT	Frequentist Probability Theory
GAISE	Guidelines for Assessment and Instruction in Statistics Education
IBME	Inquiry-Based Mathematics Education
MINEDUC	*Ministerio de Educación*: Ministry of Education, Chile
PSU	*Prueba de Selección Universitaria*: University Selection Test, Chile
SJT	Subjective Theory of Probability
UVM	*Børne- og Undervisningsministeriet*: Ministry of Education, Denmark

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Activity 1: Single coin toss.
Two players choose heads or tails. They toss a coin and whoever guesses wins.
1.1. Do you think this is a fair game? Why?

> STUDENT: The game is not fair, because the results depend on chance and not on personal abilities.

(Elicer & Carrasco, 2017, p. 751)

1.1 Motivation

During my experience as a high school mathematics teacher (2012-2013), I came across a particular discontinuity, in contrast to my background in engineering: decision making was not part of the teaching materials in the curricular units of data and chance (probability and statistics). Not as expected learning outcomes, nor as suggested exercises in textbooks. Students were expected to calculate probabilities and descriptive statistical parameters, but no course of action would follow from a result.

The missing connection between probabilities and decisions inspired my first incursion into the field of mathematics education research, once I joined a higher education institution. Supported by a colleague, we tried out paradigmatic games with students, such as coin tosses and the Monty Hall problem (Elicer & Carrasco, 2016, 2017). We aimed to explore the value of decision-making tasks as a didactical tool. Our hypothesis was, roughly, that once students realised they made *bad* decisions, a cognitive conflict may trigger a quest for an explanation. In turn, it could facilitate a better grasp on probabilistic ideas that enabled students to make *right*, rational decisions. For us, the ultimate self-evident purpose of mathematics education was to drive students away from flawed intuitions towards rational behaviour.

We picked coin tossing and the Monty Hall games, as they provide more or less predictable responses on behalf of the students. I want to focus on the unpredictable outcomes. To set the scene, we began with a simple single coin toss game, upon which a student found that "the game is not fair, because the results depend on chance and not on personal abilities" (Elicer & Carrasco, 2017, p. 751). At the time, the teachers read the student's name and his answer, rolled their eyes, and suggested to disregard it. The Monty Hall problem (e.g. Batanero, Fernandes, & Contreras, 2009) is inspired by a TV game show, where the conductor offers the player to choose one of three doors. One contains the big prize of a car, while the other two contain goats. After the player chooses, the conductor opens one of the goat doors and offers the possibility to switch. To our surprise, even after explaining the mathematics behind and empirically showing that switching to another door led to a higher probability of success, some students would stay with their first choice. They argued that the TV conductor might have a reason to want the player to switch (Elicer & Carrasco, 2016). Why would they want to help me and give away a car? These answers can easily be disregarded as being irrational or distracting, feeding the self-evident aim stated above: we need more and better probability education to produce people that do not fall into that sort of nonsense. Alternatively, one may clean the tasks and rip them out of real-world context.

Once enrolled as a PhD student, my position has evolved. Reviewing that early work, now I find it of utmost interest that, upon seemingly trivial games of chance, students reflected about the rules of the game, looking at a broader picture. I decided to go on a journey to explore and question some underlying assumptions I held. Decision making is more complicated than effective calculations. School can be a place that enables students to reflect upon broader structures shaped by data and chance. That is what this thesis is about.

1.2 Coherence along a *problématique*

As a newcomer into the field of mathematics education, I became used to hear the word *problématique* in academic conversations and adopted it without much scrutiny. Throughout the thesis, I make use of the term as a spinal cord of the research project, so I shall begin explaining my understanding of it.

A thorough discussion on the meaning and use of the term *problématique* in mathematics education can be found in Wedege (2006). She finds different connotations to the word in the literature, highlighting that what defines *the* mathematics education *problématique* is in constant evolution. Hereby, it is more appropriate to speak about *a* mathematics education *problématique*, instead. Investigative work during the first half of the 20th century was centred in the so-called *didactical triad*, constituted by the mathematics, the teacher and the student. It has only been growing in reach ever since from focusing on the mathematical content to the interconnections within. The triad as a whole unit was then contextualised in the classroom. The classroom was positioned within the school organisation, parents' background and labour market, and so on (Valero, 2010, 2012). Overall, "a mathematics education *problématique* is a historical product that can appear in response to difficulties both in the theoretical and teaching practices (in a broad sense)" (Wedege, 2006, p. 326, my translation).

Wedege (2006, p. 328) suggests that a research or practice *problématique* in mathematics education can be described as an overall and coherent quest to answer three overarching questions: (1) *what* is mathematical knowledge?, (2) *how* is mathematics learned?, and (3) *why* to teach mathematics? Niss (1993) offers a similar classification for the problems embedded in any mathematics education *problématique*. He refers to three central problem fields, namely justification, possibility and implementation, and hierarchically connects these problems. Possibilities (the *what*) can be explored once the justification question has been answered (the *why*). In turn, questions of implementation (the *how*) arise once the issues o what is possible to teach and learn according to the justifications are unfolded and given as aims.

I find that the reciprocal is also true. In order to give grounds to the claim that a conceived teaching-learning of mathematics is possible, one ought to show how is that possible, beyond our imagination. For example, if one is to claim that the inclusion of data sets from real-life contexts can be part of statistics teaching practices, it would undoubtedly help to exemplify empirically *how* it can be done. Moreover, for a justification for mathematics teaching to hold water, an underlying claim is that a mathematics teaching practice may activate the justification (Niss, 1996). For example, suppose a justification for teaching probability involves the necessity to train pupils against heuristics that lead to biases in decision making. In that case, an underlying assumption is that an education in probability *can*, in fact, have such an effect.

In their survey on *Language diversity in research on language diversity in mathematics education*, Chellougui, Thu, and Winsløw (2016), the use of the French word *problématique* is of particular relevance for the authors, where they define it, with approximations, as "a collection of *related* questions, phrased within a certain theoretical framework and, in particular, based on some fundamental assumptions related to this framework" (Chellougui et al., 2016, p. 264). I emphasise the premise that the research questions within a *problématique* not only represent parts of a whole theoretical framework, but they are supposed to be interconnected.

In this project, I pose a set of five research questions in the context of probability and statistics education in the Chilean upper secondary school realm. As Chellougui et al. (2016) suggest, these questions are interconnected, as they inform each other in matters of curriculum, design of learning environments and teaching practices. Moreover, these interrelated questions address exemplary problems of justification (why), possibilities (what) and implementation (how) of mathematics education (Niss, 1993; Wedege, 2006). In turn, these three exemplary problems are situated within a particular theoretical underpinning, namely the Critical Mathematics Education programme, which I describe next.

1.2.1 Critical mathematics education

Critical Mathematics Education (CME) can be portrayed as a research agenda grounded in a set of preoccupations, related to issues of power in mathematics and mathematics education (Skovsmose, 2014a; Valero, Andrade-Molina, & Montecino, 2015; Ernest et al., 2016). In other words, it encompasses a broad set of investigative approaches whose concerns relate to the political in technologies associated with mathematics. Its standpoint is the demystification of mathematics as intrinsically related to wonders, intrinsically related to evils, and as being neutral (Skovsmose & Valero, 2012).

According to Valero et al. (2015), three primary theoretical roots informed mathematics education researchers to give shape to CME. First, researchers drew upon developments of Marxist-inspired Critical Theory from the Frankfurt School, as a way of examining relations of power and imagine alternative possibilities to established beliefs and practices regarding mathematics education. Second, Paulo Freire's *pedagogy of the oppressed* grounded in his work as a language teacher in adult education, influenced the notion of critical education as a means for the emancipation of oppressed groups. Skovsmose (2014a, p. 116) points out that "critical education was far from expressing any interest in mathematics", thus CME led to a reinterpretation of Freire's version of *literacy* as an ability to critically read the world with mathematics. The notion of *mathemacy* enables people to confront mathematics education as motivated by ideological, political, cultural and economic interests (e.g. Skovsmose, 1998, 2005b). Third, Ubitaran D'Ambrosio's ethnomathematical perspective positioned mathematics and mathematics education as culturally situated productions, broadening the scope of mathematics learning and teaching from what was mainly seen as individual, cognitive and intellectual activities (e.g. D'Ambrosio, 1994).

These inspirations forged the so-called "social turn" in mathematics education research (Lerman, 2000), by which CME took part in exploring three overarching concerns: the notion of mathematics in action, the relation between mathematics education, democracy and social justice, and the imagination of new educational practices (Valero et al., 2015).

Mathematics in action (Skovsmose, 1994) is a reflection on the more or less explicit uses of mathematics in society and consequent effects, depriving its innocent or neutral character. Mathematics not only affects our conception of the world through its discourses and rationalities but the use of mathematical models in countless applications to technical and political decision making configures its formatting power (Skovsmose, 2005b, 2011). That is, beyond its descriptive and predictive functions, mathematics plays a prescriptive role in society (Davis & Hersh, 1986). The critical position upon this reality is that "useful things are conflated with calculable things and thus formal reasoning based on quantification, which is made possible through the use of mathematics, is purely instrumental reasoning" (Jablonka, 2014, p. 123). Therefore, CME takes part in the critique of the Enlightenment's optimism towards a scientific view of the world.

The second overarching preoccupation of CME is the non-trivial relation between mathematics education, democracy and social justice. Observing the thinking of John Dewey (Dewey, 1916/1966), Alrø and Skovsmose (2002, p. 253) criticise what they call the assumption of "harmony between scientific methodology, inquiry-based learning processes, and an endeavour for democracy". They claim that a critical perspective should also critique science and mathematics itself, i.e. it does not suffice to be a scientific or mathematical inquiry for an educational practice to be democratic. Ideals of democracy relate to self-determination or self-ruling, so a mathematics education connected to democracy ought to provide opportunities for deliberation and decision making (Skovsmose, 1998; Valero, 1999) and to be accessible to all citizens (Skovsmose & Valero, 2002). A comprehensive review of the state of the art on mathematics education and democracy can be found in Aguilar and Zavaleta (2012). The authors highlight that most of the research about this relationship is theoretical, and further empirical research needs to be done. Inspired in Freire's *pedagogy of the oppressed*, CME is meant to address different forms of suppression and exploitation (Skovsmose, 2014a). Though *social justice* is an evasive concept (Sriraman, 2007), in this line of research, CME scholars address the role of mathematics education as a – sometimes intentional – gatekeeper for social and economic opportunities (Skovsmose, 2005b). Moreover, the push for a mathematics education accessible "for all" is associated to processes of in(ex)clusion (Valero, 2017). (Valero et al., 2015, p. 13) also single out a body of research on the biases in mathematics educational achievement against certain groups. These studies, mostly from the United States, United Kingdom and Australia, focus on groups defined by race, gender, social class, migratory status, among other features.

Gutiérrez (2008) warns, however, that the "fetish" in investigating the achievement gap may perpetuate negative and static narratives about disenfranchised social groups. This research trend would counteract the very motivation for pursuing social justice.

The third major concern of CME is the imagination of new educational practices in mathematics. A critical paradigm has an ontological position of historical realism (Scotland, 2012), a viewpoint from which reality is shaped by values of a social, political, economic, cultural and ethnic character. What is actual is a crystallisation of what once was plastic (Guba & Lincoln, 1994). Upon this realisation, Skovsmose and Borba (2004) propose the notion of *pedagogical imagination*, inspired in the notion of *sociological imagination* developed by Mills (1959/2000), as a way of realising that what pedagogical practices in mathematics currently *are*, *need not to be* the case. The critical paradigm implies critique, transformation, advocacy, activism and even conflict on behalf of the inquirer Guba and Lincoln (1994), and so Skovsmose and Borba (2004) suggest a close collaboration with teachers in defining the "current situation", "imagined situation" and "arranged (or designed) situation" in educational practices. Some iconic developments include the work of Alrø and Skovsmose (2002) about the possibility to engage in *dia-logical* – as opposed to mono-logical – learning practices, guided by *intentions* that enable *critique*. A strong resonance in educational practices with the preoccupations of CME can be found in participatory action research (Atweh, 2004) and radical mathematics education (Frankenstein, 1989; Gutstein, 2006), which focus in transformation and changing students' world with mathematics.

About the latter concern, Valero and her colleagues warn about falling into the same practices a CME researcher ought to distance herself from, i.e.

> (...) there is a certain danger in technifying, as guidelines for a curriculum, the ideas of a perspective which is in itself critical of the effects of technologies associated with mathematics. (...) However, a new risk lies in increasingly taking the preoccupations of critical mathematics education as a learning and teaching theory to be "applied" in the classroom.
>
> (Valero et al., 2015, p. 15, my translation)

They refer to cases where the "gap-gaze" between social groups (Gutiérrez, 2008) has attempted to be remedied by designing mathematics curricula for the "winners" and for the "losers", perpetuating conceptions about different kinds of people. Moreover, if one is to adhere to ideals of democracy, a CME researcher must be careful. As Skovsmose and Valero (2002, p. 399) put it, "the very process of planning, carefully and in detail, access to any kind of ideas obstructs the possibility of making this access democratic". Pais, Fernandes, Matos, and Alves (2012) argue that the term "critique" has been *domesticated* and its emancipatory and reformist meanings are lost when the critical stance is confined to the capitalist school credit's system. It seems that the very nature of education systems as technologies of governance compromises the *possibility* to engage in an actual teaching-learning school practice that resonates with the preoccupations of CME.

Additional tension is recognised, for example, in the Nordic countries, as illustrated in the pre-ICME-10 document *Mathematics education – the Nordic way*. Dahl and Stedøy (2004) give a thorough revision of the parallel between democratisation processes of Nordic societies and a path to mathematics education accessible for all. In their conclusion, they point to a problem of *implementation*, as "one might then ask whether it makes sense to talk about a Nordic democratic mathematics education. As pointed out above, it is one thing to have a curriculum; classroom practise might be very different" (Dahl & Stedøy, 2004, p. 8). They recognise the challenge of offsetting centralised planning to achieve universal access with the autonomous participation of self-determined educational communities.

This PhD project is, therefore, embedded in the *problématique* of Critical Mathematics Education as an inspiration to an actual and coherent teaching-learning practice. As a way of definition, "critical mathematics education can be characterised in terms of concerns: to address social exclusion and suppression, to work for social justice in whatever form possible, to try open new possibilities for students, and to address critically mathematics in all its forms and application" (Skovsmose, 2014a, p. 116). However, scholars admit that the concerns of CME are framed by adopting evasive and explosive concepts, such as critique, reflection, and social justice (Skovsmose, 1994). Moreover, attempts to investigate and solve the issues raised by its concerns can fall into practices that contradict their ideals. Addressing the formatting power of mathematics can overestimate its role in otherwise economic, political and clientelist power structures (Valero, 1999). Access to mathematics for all can produce double gestures of inclusion and exclusions (Popkewitz, 2004; Valero, 2017), and its top-down application through the curriculum can be deemed as non-democratic (Skovsmose & Valero, 2002). Investigating and attempting to remedy educational achievement gaps in mathematics can perpetuate discourses of inferiority about social groups (Gutiérrez, 2008; Valero et al., 2015). Conceiving critique as a learning goal within the confinements of the school system can contravene the radical critique of capitalism (Pais et al., 2012).

This PhD project is an exemplary exploration of the chain of tensions that can materialise along with different problem fields of mathematics education. The thread begins with preoccupations of CME and their role in justifying the inclusion of mathematics in the school curriculum. The following issue is whether it is possible to design learning environments coherent with a critical justification within curricular boundaries. Finally, if such learning environments are deemed to be possible, tensions and challenges may emerge during the implementation of such practices in actual educational contexts. A *problématique* so defined is, of course, too overarching for a humble PhD project. In the following subsection, I describe what I focus on more specifically.

1.2.2 Delimitation process: Probability and statistics, Chile, upper secondary school

Probability and statistics

The preoccupations of CME stated above have been investigated regarding mathematics and mathematics education in general. If one conceives them as branches or applications of mathematics, probability and statistics have peculiar ties to those preoccupations. Moreover, it seems appropriate to explore the particularities of probability and statistics in the classroom, as they have increasingly been included as mainstream parts of the mathematics school curricula (Jones & Thornton, 2005).

Why probability? The historical development of probability as a mathematical discipline has ever been intertwined with the development of theories of rational choice (Mateos-Aparicio, 2002, 2004). On the one hand, decision problems have pushed the development of probability, being games of chance the most classic examples (Chernoff & Sriraman, 2014b). On the other hand, probability theories have played the role of formalising rational decision theories, including utility theory (Friedman & Savage, 1948) and game theory (Von Neumann & Morgenstern, 1944). Overall, probability plays the role of mathematising uncertainty and risks, a tool for domesticating chance (Hacking, 1990) and making decisions rationally. As Borovcnik (2015) puts it, risk and decision making define the logic of probability.

One of the preoccupations of CME aforementioned is the formatting power of mathematics. Mathematics and its models have effects that inform action. That is, "mathematics can be interpreted not only as a descriptive tool but also a source for decision making" (Skovsmose, 1998, p. 196). However, beyond the theories of individual choice with personal consequences, probability-based decisions are also made in the social and political realm in our modern risk society (Beck, 1992, 2000). Greer and Mukhopadhyay (2005) points out several of these issues, exemplifying with how probability calculations have been used to decide upon the conviction of a suspect and the probabilistic language used in the media that affects public opinion.

Moreover, one of his suggestions is precisely to conceive an "education for probabilistic modelling as a tool for critical analysis of social and political issues" (Greer & Mukhopadhyay, 2005, p. 314). Such is the case of the experience of Nobre (1989) in addressing an illegal but popular "Animal Lottery". The students in the study not only made mathematical sense of the odds, but the experience enabled them to realise the profit drive of the owners and its political ramifications in the form of corruption. Another example is the study of Savard (2015), wherein the object of research is the perception of risks in gambling by students. Beyond making sense of gambling games and pointing to the risk of losing money, at least one student pointed out the risk of addiction embedded in gambling as a social problem.

Perhaps, the formatting power of probability in society can be best encapsulated in Ian Hacking's up-front warning in his book *The Taming of Chance*:

> Ethics is in part the study of what we do. Probability cannot dictate values, but it now lies at the basis of all reasonable choice made by officials. No public decision, no risk analysis, no environmental impact, no military strategy can be conducted without decision theory couched in terms of probabilities. By covering opinion with a veneer of objectivity, we replace judgement by computation. (Hacking, 1990, p. 4)

Why statistics? For once, statistics, as we know, is intrinsically connected to probability, since chance models help statisticians make sense of variation in data (Wild, 2006; Moore, 2010). Early applications of statistics were explicitly political, as they concerned numerical information about matters of the State (Donnelly, 1998). Statistics was meant to give identity to the newly consolidated nation-states and hold things together (Desrosières, 1991). As data collections accumulated and classical theories of probability thrived, the applications of statistics as the science of the state reached almost every aspect of human life, including industry, labour, poverty, education, sanitation and crime (Porter, 1986). Nowadays, we live in a world drenched in statistical information and data-based arguments.

In terms of the preoccupations of CME, researchers have called for noting the role of mathematics in politics as a concern for mathematics teaching, leading eventually into the use of statistical information. For example, Sánchez Aguilar and Blomhøj (2010, 2016) exemplify the use of mathematical models in Mexican society, such as in the definition and use of marginalisation index, the use of mathematical models in political discourse, and the graphical representation of information. What these examples have in common is the collection and sense-making of aggregated data, i.e. a version of statistics. Some who have responded the call from other branches of mathematics eventually deal with statistics. Brantlinger (2014), for example, in the context of a geometry course realises that statistical information has to be brought to his classroom to engage in critical classroom discourse about issues of race. De Maio (2007) uses the controversial case of reporting Iraqi deaths during the American invasion in an introductory course to quantitative methods. He highlights the relevance of exploring "the role of ambiguity in statistics and the place of politics in the framing of statistical results", in order to " inspire a belief (in students) that statistics matter in the real world" (p. 37). Statistical investigations categorise kinds of people (Popkewitz & Lindblad, 2018) and operationalise social variables that are defined and reported according to the interests of governments (Frankenstein, 1994). Statistics is a mathematical technology entrenched in issues of power.

My view about statistics is that not only can one find compelling examples of its concepts and procedures applied to social and political issues. To some extent, the political effects are at the core of statistics as a subject-matter, as its original coinage suggests: the arithmetic of the State (Achenwall, 1749). Furthermore, statistics has a formatting power in society, since data, variation, and chance are omnipresent in modern life (Moore, 1998).

Chile

The empirical basis of the thesis is situated in Chile, my home country, and there are some pragmatic reasons for this choice. First, the collaborative work with teachers I had planned for was more fluid, in terms of familiarity with the educational context and common language. Moreover, since the plan is to include contingent cases into the classroom experiences, it is much easier to do in my own country. Additionally, as indicated in the acknowledgements, this PhD project is funded by the Chilean National Agency for Research and Development (ANID), under the programme *Becas Chile*. As part of the grant agreement, I must return to the country and reside for a determined number of years as retribution. In anticipation, I wanted my thesis to be of interest to the Chilean research and educational communities, to be a contribution. My future in research and teaching is there.

Another element I did not plan for is the convergence of two national events. On the one hand, a political spark in October 2019 triggered social unrest and massive protests, which, in turn, have led to spontaneous town halls and a national debate about the country we would like to become. As a result, in October 2020 Chile will hold a referendum to decide whether the people will initiate the writing of a new constitution and abandon the one installed during the dictatorship in 1980 and still prevailing. On the other hand, new curricular bases for upper secondary school have just come out of the oven (MINEDUC, 2019) and will begin to be implemented as soon as the current pandemic allows for it. This document promotes greater attention to democratic values, citizenship and interdisciplinary investigative work. I believe some of the outcomes and reflections in my thesis echo the historical moment Chile is going through, one of a society rethinking itself.

Upper secondary school

My motivation for enrolling into a PhD programme comes from my experience as a high school teacher in mathematics in Chile. To be clear, in the Chilean context, I refer to high school as the last four years of secondary school, namely grades 9th to 12th, with students aged from 14 to 18 years old. In the last decade, the general curricular bases divide the secondary school into two blocks: from 7th to 10th grades, and from 11th to 12th grade. I narrow down the scope of the thesis to upper secondary school, i.e. the last two years of high school education as a curricular unit explicitly more dedicated to preparing students to become citizens (MINEDUC, 2017). Given the focus on the connections between mathematics education and citizenship in the CME agenda (Skovsmose, 1998; Valero, 1999), I would like to aim attention to the latest mandatory educational stage before adulthood. For many students, the upper secondary school may be their last formal encounter with mathematics. Therefore, if there is an opportunity to make a smoother connection between mathematics and the out-of-school critical citizen life, this is it.

1.3 The research project: What did I do?

As a way of encompassing and organising mathematics education as a field of research, Valero (2010, 2012) suggests to visualise it as a network of social practices, not only accounting for the widening of objects of research (individual, family, school, national and international arenas) but the complex interconnections therein. Though such complexity is broadly acknowledged, Valero argues that, for several reasons (financial included), objects of study are examined isolated from others. As a consequence, many studies show that such limited approaches do not solve problems in their complexity. Furthermore, she proposes at least three strategies for future research taking this complexity into account; first, to study some of the interconnected practices; second, to move between micro and macro structures; and third, to explore a particular constellation throughout time. This research project can be exemplary for the second strategy, that is, it makes sense of the movement from the macro- to the micro-social and back (de Abreu, 2000). The overall PhD research project is decomposed into three parts, where justifications, learning environment designs and teaching-learning practice are the corresponding objects of research. This section can be seen as a depiction of the process timeline, as illustrated on Figure 1.1.

The first part of the project is an exploration of notions of risk and decision making at a macro-level, both embedded in the historical development of probability and statistics as scholarly knowledge and educational research, mostly influenced by the psychological, educational agenda of heuristics and biases. From here, I intend to move towards the micro-level. Therefore, the core of this study is an exploration of the role of decision making in the curriculum. That is, regardless of scholarly justifications, official curricular documents represent a concrete connection between the subject matter and actual educational practice carried out by teachers. I use a version of Critical Discourse Analysis (Fairclough, 1992, 2003, 2010) to reveal how mentions of decision making legitimise the inclusion of probability and statistics in Chilean and Danish upper secondary school curricula. This study led to two publications (Elicer, 2019, 2020). The following research questions guide the latter and most comprehensive of them:

RQ1 What is the role given to notions of decision making in the Chilean and Danish upper secondary school probability and statistics curricula?

RQ2 Which similarities and differences can be drawn from these two cases?

At this stage of the project, no mentions to CME are presumed, since through **RQ1** and **RQ2** I let curricular texts to speak. The study allows me to show different justifications entrenched with notions of decision making, some of which seem compatible with the critical values of participation and democracy, tied to socially and politically relevant issues.

The second part of the project, connected mostly to the implementation problem, can be seen as an *intermezzo* between the teaching purposes put forward in the first part and the actual practice of the third. Here, I bring together relevant research literature and the official curricular framework to define design principles that conform to the critical justification delineated in the first part, while respecting the constraints of actual teaching practice. According to these principles, I construct a collection of inquiries that turn into learning environments once proposed to the teachers. The proposals are adapted in close collaboration with them, and thus applicable to their real teaching context. The research question guiding the second part of the study is the following:

RQ3 What considerations should there be accounted for to design learning environments compatible with a critical justification for probability and statistics, under upper secondary teaching conditions in Chile?

It must be stressed that a response to **RQ3** is speculative by merely proposing a set of design principles. A revision of them comes after the final part of the project. The goal is to acknowledge some of the tensions and challenges involved.

The third part of the project is devoted to both the possibility and implementation problems, and it comprises four classroom experiences taking place in Chilean upper secondary school classes, based on the learning environments yielded by the second part. At this stage, the goal is to uncover whether, to what extent and how the critical justification is coherent in the learning situations. The students' work and reflections during the experiences become the units of analysis. In the spirit of close collaboration, besides taking part in the design and implementation of the learning environments, two of the teachers participate in the analysis and discussion through post-intervention interviews on the following year. This stage of the research project does not only attempt to expose what was evoked during the experiences but also what potential they convey for future research and practice. Other than revising the considerations addressed in **RQ3**, I examine the classroom experiences addressing two other research questions:

RQ4 Can students reflect upon the formatting power of mathematics in learning environments designed after the considerations addressed in **RQ3**?

RQ5 What role do probabilistic and statistical notions play in students' critical reflections unveiled in **RQ4**?

RQ4 is formulated based on the critical epistemology of Skovsmose (1994); Alrø and Skovsmose (2002), i.e. a version of critique that situates the first preoccupation of CME – the formatting power of mathematics – into the classroom. The relevance of **RQ4** within the possibility problem is that whether students are experiencing the preoccupations of CME in the classroom will depend on how one defines and operationalises the meaning of critique in learning. The importance of **RQ5** is that it allows me to come back to the general *problématique* focusing on aspects of the formatting power of mathematics that are exclusive to probability and statistics as subject matters. Addressing **R5** means narrowing down **R4**, assuming that, even if the learning environments of this thesis *did* or *could* provoke critique, they could still be referring to the more general use of mathematical models and applications. These five research questions are interconnected in a sequential and subordinated manner. Findings in **RQ1** allow for comparisons in **RQ2**. These questions let me make sense of the purposes and justifications that inform **RQ3**, whose realisation in teaching practice configures the analyses for **RQ4** and **RQ5**. By travelling back through these research questions, the coherence along the general *problématique* can be addressed. The analyses of the classroom experiences will reveal some of the tensions and challenges in the implementation of teaching practice as well. These tensions can shed light on whether and what it is at all possible to teach and learn, regarding the preoccupations of CME connected to probability and statistics as school subjects. The cycle is completed by coming back to the justification problem. The critical justification for upper secondary school probability and statistics exists both in research and as a guideline for practice. We can make sense of it, provided the possibility of designing and implementing teaching situations with a so-declared purpose, and empirical evidence of students engagement and reflections.

The structure of the project is described chronologically to make sense of the work done. During the entire process of research, my points of view changed, my priorities switched, and lessons were learned. The thesis, on the other hand, is constructed in order to make sense of the findings logically and progressively, which I describe next.

1.4 General structure of the thesis: What will you read?

This doctoral thesis is written as a monograph, as opposed to a collection of scientific papers. The list of research outputs in the front matter is an account for the work done during my studies, but the dissertation is a self-contained document. The choice for a monograph is the result of an early discussion with my supervisor, agreeing on the rationale that a full dedication to a single project for four years may well be a once-in-a-lifetime opportunity. In principle, it would be helpful to make sense of my research on a single red thread throughout the thesis. I do not mean to underestimate the value of writing and publishing papers. The point is that, as an academic, the moment to do so in a systematic fashion will eventually arrive.

In this dissertation, I explore the general *problématique* employing four classroom intervention studies, which configure the core of the project. These studies aim to illustrate challenges

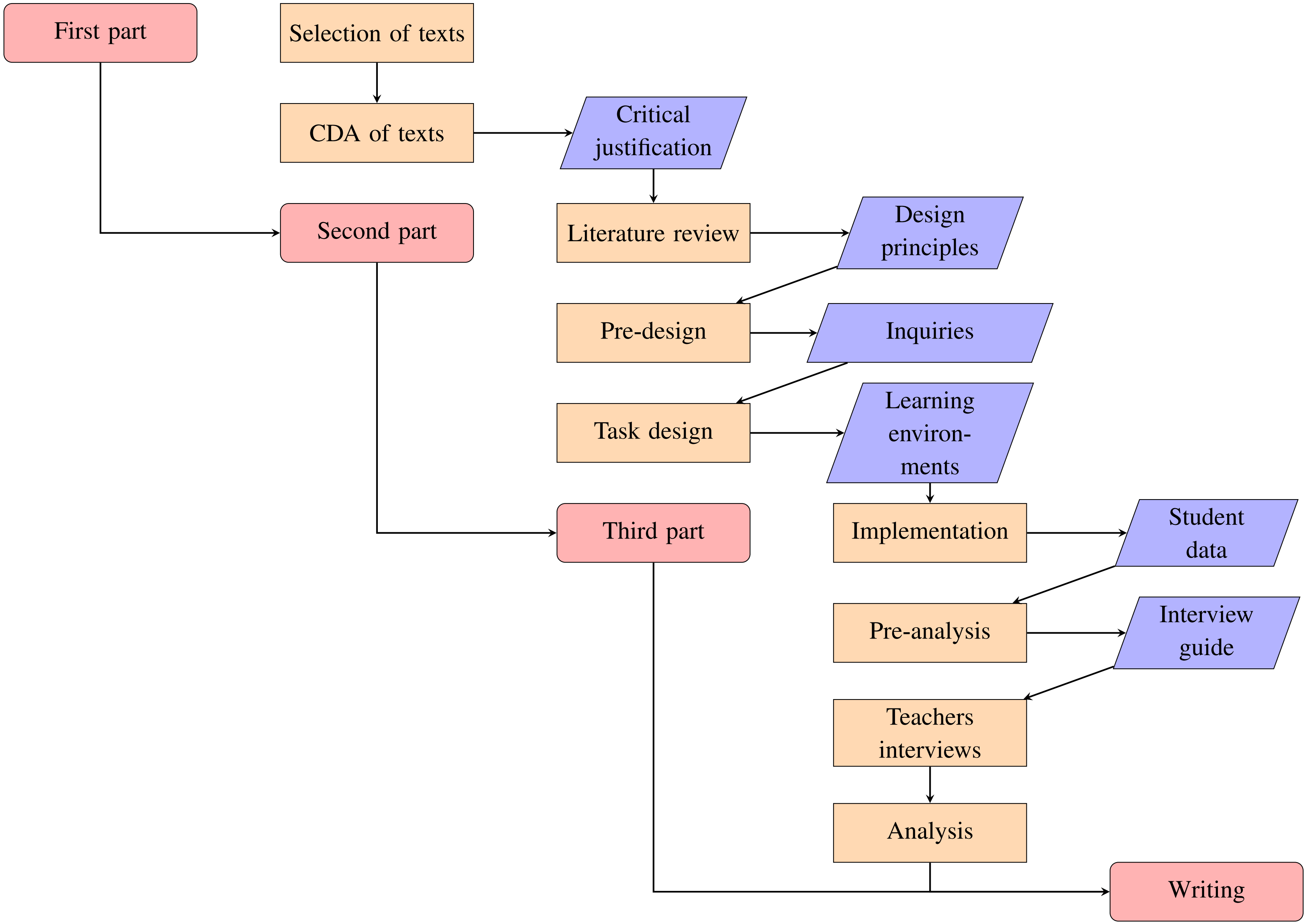


Figure 1.1: General structure of the research project

in the design, practice and research on probability and statistics education from a critical perspective. However, previous theoretical and methodological clarifications are needed.

In Chapter 2, I describe the conceptual framework for the thesis. Elements of the philosophy of CME work as a set of preoccupations and terminology that inform decisions concerning the use of other frameworks. In particular, the notion of formatting power of mathematics and the inquiry cooperation model of classroom interaction are unfolded. Critical Discourse Analysis (CDA) provides a coherent methodology to address the justification problem as mirrored in exemplary texts extracted from official curricular documents. Inquiry-based mathematics education (IBME) plays the role of a framework for action, informing the structure of learning environments and guiding the how-to in practice. Finally, different frameworks are used for the analysis of students' work during classroom interventions. These include Skovsmose's entry points on reflective knowing to address **RQ4**, and notions of probabilistic and statistical literacy to address the domain-specificity of **RQ5**.

The first link in the thesis' chain of coherence is that of justifications. In Chapter 3, I begin by exploring the notion of decision making as embedded in the history and educational research on probability and statistics. Then, using a version of Fairclough's CDA, I show that and how the notion of decision making plays a role in legitimising probability and statistics in curricular documents, taking Chile and Denmark as exemplary cases. As a result, I put in evidence that probability and statistics are claimed to be necessary and relevant beyond the personal psychological scope. Mentions to decision making open up to the social and political realms, partly connected to a critical framing of justifications. However, there is no apparent connection to the importance of learning probability and statistics as school subjects. This finding leads us to the problem of imagining a teaching practice that is oriented accordingly.

Though I address the justification problem through an analytical strategy based on CDA, the methodological approach is different throughout the rest of the thesis. In Chapter 4, I describe the more overarching approach to address the coherence between the aims depicted by the critical justification and teaching practice. First, I define and discuss three main design principles – exemplarity, inquiry approach and pragmatism – and describe how they inform the four learning environments. Second, I describe the practicalities concerning the classroom interventions that took place and how data were collected in video and audio format. Third, I dedicate a section to describe post-intervention semi-structured interviews with the teachers involved in the project. These interviews have the double purpose of triangulating part of the analysis of students' work and reflecting on possibilities that the experience opens in future practice. Finally, I describe how the analysis of students and teachers' reflections is performed, according to the frameworks described in the previous chapter.

The following chapters are titled after the inquiries that guide each of the four learning environments implemented. They are organised in such a way that the reader can follow the classroom experience in chronological order. However, the core of each section is not the activity itself, but the broader issues that are illustrated by the mathematical content, students' work and reflections, and the teachers' input.

Chapter 5 explores the notions of data and probabilistic evidence, and their connections to the principle of evidence beyond all reasonable doubt. Students are challenged to design and carry out an experiment. They use their results to compute a version of a P-value, based on the binomial model. Later, they discuss upon a real case of a wrongful conviction resulting from a problematic application of this logic. Difficulties in probability calculations suggest that adaptations are needed. The disconnection between the case study and students' daily life made the reflection hard to transfer from the hands-on experiment and probability calculations. The need to crystallise mathematical ideas and to anchor the inquiry to a more relatable event leads to the next chapter.

PSU is the battery of standardised tests undertaken by high school graduates in Chile, as an input for their applications to higher education institutions. In Chapter 6, 12th-grade students, soon to confront the test, engage in understanding how these scores are computed using a normalised relative scale. Students make sense of percentiles by relating them to how they see themselves as belonging to lower-income quintiles. Conversations on prospects for their future reveal the problematic relation between socioeconomic background and academic foreground. Understanding the relative scale allows them to illustrate how the system is designed to leave a portion out of higher education. However, the construction of the scale through a normal distribution is not made available for students. Nonetheless, they deconstruct the display of graphical representations of PSU-score trends as a political tool. Critical reflections are enabled by letting students make their version of diagrams, that is, by becoming producers and not only consumers of statistical messages. This active character of students' work reinforces the formulation of the inquiries that guide the following chapters.

The learning environment described in Chapter 7 has students assessing and deciding whether a new thermoelectric power plant should be installed in their province, using fictitious provided data. The contrast between their calculations and the results provided by the company opens up the possibility to discuss the theoretical character of independence. Questioning data sources enables students to reflect on the theoretical character of probability. Moreover, by having to make a decision, the complexity and implications of the notion of risk emerge and can be connected to broader issues of the risk society. An event that students can relate to enables these reflections, as it concerns their community. The next step is to explore an inquiry related to an event in which students had participated.

In Chapter 8, students are asked to estimate the crowd size of a demonstration some of them organised and partook in. The omnipresence of variation in statistics is illustrated by realising how different social actors provide different estimates, including their own, turning the statistic-or-lie duality into a calculation-and-construction bundle. Through statistical estimation, students recognise and negotiate the assumptions to find a credible estimate. The discussion about crowd sizes is exemplary for the formatting power of mathematics, namely the power to validate or invalidate a cause on the grounds on its apparent adherence by the masses. This reflection is enriched by the realisation of how questionable are the sources of data based on students' context knowledge, invalidating the mathematical approach proposed.

Though each classroom experience is exemplary for particular topics, some issues of research, teaching practice transcend all four of them. In an attempt to bind the experiences together in a thorough discussion, in Chapter 9 I extract fundamental lessons regarding three aspects. First, I revise the way the learning environments illustrate the formatting power of probability and statistics, as well as other preoccupations of CME. Second, I review how critical reflections are embedded in the different frameworks for analysis. Third, I delineate the principal tensions between the design principles.

To give closure to the thesis, in Chapter 10 I summarise the main findings and reflections on the research process. First, I explicitly address each of the research questions. Finally, I connect these research questions to the problems of justification, possibilities and implementation embedded in the general *problématique*.

Chapter 2

Theoretical underpinnings

The opening act to the First South American PME Conference was a plenary lecture by professor Merrilyn Goos. She reported a review of research on understanding and promoting mathematical thinking (the theme of the conference, Gómez (2018)), as portrayed in publications on *Educational Studies in Mathematics*, during her tenure as editor-in-chief (2014-2018). Goos gave a concise but thorough presentation on the context and features of different studies, making sense of methodological approaches, as well as identifying contributions to the body of knowledge and future directions. The one slide for the theme of *theories* was a messy collage of conceptual maps and diagrams, which amused the audience.

The point she wanted to make is that different from other themes of the review, finding patterns or putting theories into boxes was extremely difficult, recognising that the proliferation of theories is a broad challenge in mathematics education research (Prediger, Bikner-Ahsbahs, & Arzarello, 2008). She observes:

> Although it was possible to recognise broader families of theories emanating from cognitive or constructivist or sociocultural standpoints, it seemed that the theories in use were developed for specific *purposes* and often *combined* with other specific theories in order to illuminate a particular phenomenon.
>
> (Goos, 2018, p. 9, emphasis added)

As a PhD student, this somewhat anecdotal episode was quite illuminating. Navigating in discussions of theory in mathematics education can be a stormy business. However, a two-hand firm hold on the rudder can be an explicit declaration on the purpose of theories and a realisation that often these can be combined.

Exploring the general *problématique* of this thesis implies addressing five research questions within different problem fields of mathematics education. Accordingly, I make use of different theoretical underpinnings to establish a varied but interconnected conceptual framework.

2.1 Theories in mathematics education

According to Niss (2007a, p. 98), in a broad sense, a theory is a system of concepts and claims, characterised by the following properties:

- The theory consists of an *organised network of concepts* (including ideas, notions, distinctions, terms, etc.) and *claims* about some extensive domain, or a class of domains, of objects, situations and phenomena.
- In the theory, the *concepts are linked in a connected hierarchy* (oftentimes of a logical or proto-logical nature), in which a certain set of concepts, taken to be basic, are used as building blocks in the formation of the other concepts in the hierarchy.
- In the theory, the *claims are either* basic hypotheses, assumptions, or axioms, taken as *fundamental* (i.e. not subject to discussion within the boundaries of the theory itself), or statements obtained from the fundamental claims by means of *formal or material* (by "material" we mean experiential or experimental) *derivation* (including reasoning).

Theories serve a variety of purposes, which are not mutually exclusive. Niss (2007a) identifies six. First, a theory can provide *explanation* to some observed phenomenon within the domain of the same theory, if it can be derived from its claims. Second, a theory can give predictions for the possible occurrence of phenomena as a possible claim within the theory, based on the fulfilment of certain conditions. Third, a theory can be of *guidance for action or behaviour*, by employing knowledge of claims, in order to achieve desired outcomes and avoid undesirable ones. Fourth, a theory can provide a *structured set of lenses* to approach, observe, study, analyse or interpret particular aspects of the world. A fifth purpose is to provide *a safeguard against unscientific approaches* to a problem, by articulating underlying assumptions, choices, a consistent terminology and situating a study within a framework and body of research. The sixth purpose is to *protect against attacks* from sceptical or hostile colleagues in other disciplines, by making theoretical underpinnings explicit.

In mathematics education scholarship, despite some attempts (i.e. by Steiner (1985)) there is no unified grand *theory of mathematics education* (Lester, 2005; Niss, 2007a, 2007b; Sriraman & Nardi, 2013), and part of it is due to the complexity and scope of phenomena addressed in the field. Consequently, mathematics education research borrows theoretical underpinnings form other fields, be them epistemology and sociology of mathematics, educational philosophies, psychological theories of learning, pedagogical theories, neuroscience, linguistics, and theories from sociocultural sciences (Niss, 2007a, 2007b).

According to the scope of reach of theories, Silver and Herbst (2007) distinguish between grand, middle-range and local theories. *Grand theories* of mathematics education attempt to organise the whole field, *middle-range theories* identify specific methodologies according to the sub-field of study, and *local theories* work as a mediator in a triad of problems, research and practices. For them, as attractive as the development of grand theories could be, "more fertile ground is likely to be ground if we till the soil for middle-range theories" (Silver & Herbst, 2007, p. 61), as they can produce construct that help to visualise issues that are invisible otherwise. On that note, Niss (2007a) identifies within these middle-range theories many homegrown theories in mathematics education, including the work of (Skovsmose, 1994) towards a philosophy of critical mathematics education, which is at the core of this thesis.

2.1.1 Theories in probability and statistics education

Probability and statistics education research is no exception to the diversity and origins of theories. Jones and Thornton (2005) provide a historical account on the field of research in teaching and learning probability in school, where they define three main periods, more or less distinguished by the theoretical underpinnings drawn. The first period (circa the 1950s to 1960s) was highly influenced by the work of Piaget and Indhelder and focused on "developmental growth and structure of people's probabilistic thinking" (p. 65). The second period (circa the 1970s to 1980s), characterised by the work of Fischbein (e.g. Fischbein, 1975) and related to the work by Kahneman and Tversky (e.g. Tversky & Kahneman, 1974), focused in probabilistic intuitions and led to the so-called heuristics and biases programme (more in Section 3.2 of this document). These two phases, referred to as Piagetian and post-Piagetian (Jones & Thornton, 2005; Batanero & Borovcnik, 2016), are highly grounded in psychological theories of learning, thinking and judgement. The following period, starting in the 1990s is characterised by "the emergence of probability and statistics as a mainstream strand" (Jones & Thornton, 2005, p. 79) in the mathematics curriculum, and thus expanding the scope of research to curriculum, thinking, reasoning, and different aspects of school learning environments. Chernoff and Sriraman (2014a) argue that this period of Contemporary Research (contemporary for Jones and Thornton (2005)) "can also be historically judged for the following phenomenon: liberal utilisation of research topics in the field of mathematics education" (p. xvi). In turn, it borrowed the complexity of uses and roles of theories in mathematics education as portrayed by Niss (2007a, 2007b). Reviews on research and development in probability education such as in Borovcnik and Kapadia (2009) refer to theories only as theories of probability (e.g. classical and Bayesian), i.e. from an epistemological philosophy of the discipline. According Chernoff and Sriraman (2014a), after the Contemporary phase, there ought to be a new Assimilation period, attempting to conciliate different interpretations of probability, the interest in heuristics, the field's psychological roots and issues of risk (Borovcnik & Kapadia, 2014). However, the coinage and definition of this Assimilation period are, in Chernoff and Sriraman's words, still premature, which Batanero and Borovcnik (2016) cosign in their recent topic survey. That is, even if considered a sub-field of mathematics education, probability education scholarship shares the impetus and struggle of assimilating a unifying theory of probability education.

Though evident connections (historical and epistemological) can be drawn between probability and statistics, statistics as a discipline has a more substantial and necessary grip on non-mathematical context (G. W. Cobb & Moore, 1997; Moore & Cobb, 2000). Moreover, its interdisciplinary nature can be traced back to its original purposes and mirrored in the conformation of early statistics scholar societies (Wild, Utts, & Horton, 2018). Statistics education research prides itself to "mature as a discipline distinct from mathematics education, creating new perspectives on the teaching and learning of statistics" (Groth, 2015, p. 4), perhaps most acknowledged by the release of the first *International Handbook of Research in Statistics Education* (Ben-Zvi, Makar, & Garfield, 2018).

In a recent review, Nilsson, Schindler, and Bakker (2018) investigate and classify the nature and use of theories in statistics education research. Though they use a more flexible notion of theories as "theoretical approaches" or frameworks (p. 368), they find four overarching categories of theories informing research. The first category, namely Theories of Statistics, is of a disciplinary nature, i.e. the nature of statistics as a subject matter. It is subdivided into product-oriented (i.e. statistical big ideas) and process-oriented theories (i.e. steps in statistical investigation cycles). The second category corresponds to Theories with a Didactical Focus, focused on teaching and learning, fitting as frameworks for action and domain-specific instructional theories (diSessa & Cobb, 2004). Inevitably, the third category is that of Theories from Mathematics or Science Education, grounded in the inherent connection of statistics with these disciplines. Finally, the last category comprises Theories with a Broader Range on Epistemological Aspects, such as Vygotsky's learning theory and Bordieu's understanding of culture. Moreover, the authors address a second research question about the extent to which theories are combined. They conclude not only that theories of different types are extensively combined. They find "drawing on multiple theories promising because, in our experience, different theoretical resources are typically needed to study complex issues in depth (cf. diSessa & Cobb, 2004)" (Nilsson et al., 2018, p. 380).

Theories in mathematics education are systems of interconnected concepts and claims that attempt to depict, organise, explain and inform different aspects of mathematics, learning, teaching and implications. There is no such thing as a unified "theory of mathematics education". However, some theories are more overarching than others and often are borrowed from other fields and combined, playing different roles (diSessa & Cobb, 2004). The case is no different in probability and statistics education. The review of theories in statistics education by Nilsson et al. (2018) is exemplary for yet another pinpoint. In their search, they use an open-minded approach to explore the selected papers, given that theories are not always explicitly separated from the method or the formulation of research questions. As a result, what counts as theory sometimes is referred to as theoretical underpinnings or theoretical framework, which deserves some clarification.

2.2 Frameworks

One difficulty I have found in the literature regarding theories in mathematics education is the distinction between the notion of theories and frameworks. For example, within the realm of Design-Based Research (DBR) (P. Cobb, McClain, & Gravemeijer, 2003; Design-Based Research Collective, 2003), given its dual goal of developing theories of learning and the means to support it, researchers make use of theories (or frameworks) of different levels. In their PhD-friendly article, Bakker and van Eerde (2015, p. 437) provide the following guide:

> To address the role of theory in DBR, it is helpful to summarise diSessa and Cobb's (2004) categorisation of different types of *theories* involved in educational research. They distinguish:
>
> - Grand *theories* (e.g., Piaget's phases of intellectual development; Skinner's behaviourism)
> - Orienting *frameworks* (e.g., constructivism, semiotics, sociocultural *theories*)

- *Frameworks* for action (e.g., designing for learning, Realistic Mathematics Education)
- Domain-specific *theories* (e.g., how to teach density or sampling)
- Hypothetical Learning Trajectories (Simon, 1995) or didactical scenarios (Lijnse, 1995; Lijnse & Klaassen, 2004) formulated for specific teaching experiments (...)

I have added emphases to highlight the fact that, in their formulation, *theories* and *frameworks* are depicted as almost interchangeable terms. At least, it appears to be that some frameworks are types of theories. In his development of a theory of authentic task situations, Palm (2009, p. 6) observes that, based on the ideas of Niss (2007b) and Lester (2005) "both notions include an organisation of concepts and ideas, but that a framework does not need to include claims about phenomena. Thus, a theory may sometimes be thought of as including a framework and claims". For him, frameworks define the conceptual elements of theories.

On the other hand, Eisenhart (1991) classifies research *frameworks* into three categories: *theoretical*, practical and conceptual. She draws on a general view of frameworks as skeletal structures to support and enclose. Theoretical frameworks guide research relying on formal theories providing well established and coherent explanations to certain phenomena. She exemplifies Piaget's theory of conservation, Vygotsky's theory of socio-historical constructivism and Simon's theory of human problem-solving as such formal theories. Practical frameworks arise as an alternative for practitioners, who often do not find theories to work for them. "This kind of framework is not informed by formal theory, but by the accumulated practice knowledge of practitioners and administrators, the findings of previous research, and often the viewpoints offered by public opinion" (Lester, 2005, p. 459). Finally, conceptual frameworks are "skeletal structures for justification, rather than explanations" (Eisenhart, 1991, p. 210). The key point is that a conceptual framework consists of an argument for choosing specific concepts and their interrelations in light of the purposes of research, and it can be informed both by theories and practice. Eisenhart (1991) and Lester (2005), therefore, do not portray theories and frameworks as interchangeable. Instead, depending on the purposes of a research project, it can be *framed* in different ways, and theories, as well as research literature and accumulated practice knowledge, can aid the choice and scrutiny of such frameworks.

From this perspective, I find the most honest framing of my research project to be through a conceptual framework, arguing for the choice of different concepts and their interrelations. Some of these concepts are rooted in theories. For example, the critical epistemology of mathematics (Alrø & Skovsmose, 2002) is inspired in both the constructivism of Piaget (1970) and sociocultural approach of Vygotsky (1978). Some frameworks are drawn from the literature, such as different perspectives of probability and statistical literacies (e.g. Gal, 2002, 2005). Some frameworks are grounded in practical issues, such as the principle of pragmatism for the design of learning environments.

2.3 Conceptual framework

In Chapter 1, I have characterised Critical Mathematics Education as a set of preoccupations within a broad body of research, focusing on issues of political and societal relevance connected to mathematics education. Several concepts take part in these concerns (e.g. critique, social justice, pedagogical imagination), as well as claims (e.g. mathematics is not intrinsically good, nor bad, nor neutral). However, I am not aware of scholars naming CME as a theory in itself following the definition by (Niss, 2007a, 2007b). In part, some of the concepts put forward are admittedly too evasive or explosive Skovsmose (1994); Alrø and Skovsmose (2002), in the sense that each definition opens up to broader, more complicated concepts. For example, the *critical* aspect of CME could refer to the awareness of being in a critical or turning point and convey action towards that *crisis*. It can yield value-based or normative *criticism* of mathematics, society and education. *Critique* can also be an analytical evaluation of hidden social and cultural dimensions in mathematics and mathematics education (Ernest, 2010). However, "critique cannot be any dogmatic exercise, in the sense that it can be based on any well-defined foundation. In particular one cannot assume any specific interpretation of (...) critique (...) We have to do with concepts under construction" (Skovsmose, 2014a, p. 119). The way to approach and investigate the preoccupations is called by CME scholars to admit no dogmas and embrace uncertainty (Skovsmose, 2014b; Ernest et al., 2016).

Nonetheless, for the sake of moving on, a crucial notion of CME throughout the thesis is the *formatting power of mathematics*. Niss (1994) delineates five perspectives from which mathematics can be seen as a discipline, namely as pure science, applied science, a system of instruments, a field of aesthetics and school-subject. The formatting power of mathematics refers to the third perspective, which Niss (1994, p. 367) unfolds as follows:

> Mathematics is also a system of instruments, products as well as processes, that can assist decisions and actions related to the mastering of extra-mathematical practice areas. (...) This mathematics provides tools for the exercise of a very wide range of social practices and techniques.

These mathematical systems become crystallised as part of the normal functioning of an increasingly automated society. Moreover, these systems are often obscured, whether intentionally or as a consequence of unintended elitist access to mathematical knowledge (Boaler, 1997; Skovsmose & Valero, 2002).

Mathematics having a formatting power does not mean that mathematics is omnipotent, but that it plays a role among other tools and rationalities. Skovsmose clarifies:

> Formatting was originally described, in abstract terms, as a path leading from thinking abstractions to realised abstractions. This terminology provides an idealistic formulation of the potential of mathematics and its effects.
>
> ...
>
> I do not point to any interpretation of mathematics as being the 'prime mover' of society. We have to recognise an interplay, and we always have to take into consideration different social forces ... An interaction takes place. Let this be a meaning of 'mathematics is formatting our society' " (Skovsmose, 1994, pp. 136–137)

Another concept from CME I use throughout the thesis is focused on classroom interactions. Alrø and Skovsmose (2002) delineate elements of dialogue for communicating and negotiating intentions, fostering reflections and encouraging critique in classroom activity, in what they denominate the inquiry-cooperation model. This model consists of a set of eight speech acts that take place between students and with the teacher, namely getting in contact, locating, identifying, advocating, thinking aloud, reformulating, challenging and evaluating.

"*Getting in contact* means tuning in to each other in order to prepare for co-operation" (p. 62). It can be achieved employing inquiring questions, tag questions, mutual confirmation, support and the use of humour. *Locating* "means finding out something that you did not know or was not aware of before" (p. 101). It becomes evident when examining possibilities and trying things out. *Identifying* means putting what is located in mathematical terms. Through justifying findings and further crystallisation of mathematical ideas, "the process of identification will provide a resource for further inquiry" (p. 62). "*Advocating* means putting forward ideas or points of view not as the absolute truth, but something that can be examined". It is manifest when a position is taken, which can be confronted. "*Thinking aloud* means expressing one's thoughts ideas and feelings during the process of inquiry" (p. 107). "*Reformulating* means repeating what has just been said, maybe in slightly different words or tone of voice" (p. 108). It can be done by paraphrasing or by completing utterances. "*Challenging* means the attempt to push things in a new direction or to question already gained knowledge or fixed perspectives" (p. 109). Asking hypothetical questions is a commonly used resource to challenge students and bring the inquiry to a turning point. *Evaluating* can be "correction of mistakes, negative critique, constructive critique, advice, unconditional support, praise or new examination - the list is incomplete" (p. 109).

The influence or possible effects of dialogical features are out of the scope of the project. For the sake of this thesis, the inquiry-cooperation model provides a consistent terminology for understanding (not explaining) interactions in the classroom experiences, i.e. "a language of description of an educational practice" (Silver & Herbst, 2007, p. 56).

2.3.1 Critical discourse analysis

The first exploration in the thesis concerns the role that decision-making plays in curricula. It began as an assignment for a PhD course and turned into two papers (Elicer, 2019, 2020). The research questions given in the introduction as **RQ1** and **RQ2** do not specify yet what roles are and how to observe them in data. I draw on one possibility Silver and Herbst (2007) give to the use of a "theory as a means to transform a commonsensical problem into a researchable problem" (p. 50).

I draw on Fairclough's three-dimensional framework for Critical Discourse Analysis (CDA) (e.g. Fairclough, 1992; P. Cobb et al., 2003) to position texts within a broader scope. The broader dimension is constituted by *social structures* and define what is at all possible (e.g. language). The objects of research are *social events* and represent what is actual and observable (e.g. texts). Discourse, Fairclough (2010, p. 230) recognises, "is commonly used in various senses, including (a) meaning-making as an element of the social process, (b) the language associated with a particular social field or practice (...), and (c) a way of construing aspects of the world associated with a particular social perspective". I draw on the first interpretation, namely discourse as a process of meaning-making or semiosis. I am attempting to understand the meaning or function of selected texts.

From this perspective, texts are deemed to be concrete examples of discursive practices. Given that the research questions concerning curricula are embedded in the justification problem (Chapter 3), amongst different discursive practices to spot, I focus on legitimation strategies (van Leeuwen, 2007) and the positioning of social actors (Fairclough, 2003). The point is, first, to see whether and how mentions to decisions legitimise probability and statistics in the curriculum and, second, whom these decisions are referred to.

The role of CDA is to refine the research questions and thus provide a coherent methodology. First, at a textual level, I analyse according to different choices in the use of modalities, collocations, and intertextualities (Fairclough, 2003). Then, at a discourse level, the "roles" of decision-making are the way decisions take part of discursive practices (positioning of social actors and legitimation) as illustrated in the texts. Using the characterisation of (Niss, 2007a), the purpose of theory is to provide a structured set of lenses to approach, observe, study, analyse and interpret texts as a particular part of the world. Moreover, in the process of publishing Elicer (2020), one of the reviewers suggested blending what was under the headings of *theory* and *methodology*, into an overall *analytical strategy*, as can be seen in Section 3.3. After all, at least in this part of the research project, "we cannot so sharply separate theory and method" (Fairclough, 2010, p. 234). This layout is the main reason why, in order to avoid redundancy, I do not refer back to CDA in Chapter 4 (Methodological generalities) after Chapter 3 (The justification problem).

2.3.2 Framework for action

Speaking of not separating theory and method, the second part of the project consists of designing learning environments that are compatible with the critical justification for teaching and learning probability and statistics in upper secondary school. That is, a frame is needed to address **RQ3**. Hereby, educational philosophies can give *guidance for action or behaviour* (Niss, 2007a) in the form of *frameworks for action*, in the classification of the role of theories of diSessa and Cobb (2004).

Skovsmose (1994) draws on Dewey's philosophy of experiential learning to argue for a type of learning practice compatible with the ideals of democracy and empowerment. Dewey attempts to disrupt traditional top-down educational practices and put students' experiences at the core of the learning process. As (Hiebert et al., 1996, p. 14) put it, Dewey "believed reflective inquiry was the key to moving beyond the distinction between knowing and doing". Still, Skovsmose is ambivalent about Dewey's philosophy, in his optimism towards the scientific method as a teaching-learning practice (Skovsmose, 1998). On the one hand, he agrees that education and democracy are aspects of the same discussion. On the other hand, Dewey is not critical of the scientific method, which he sees as "the only authentic means at our command for getting at the significance of our everyday experiences of the world in which we live" (Dewey, 1938/2015). That is, an "education which organises itself in line with an inquiry process becomes 'education for democracy' " (Skovsmose, 1998, p. 196) as a necessary – but not sufficient – condition.

Skovsmose proposes a two-dimensional overview of possible ways of organising learning environments, summarised in Table 2.1 (Skovsmose, 2011, p. 40). On one dimension, mathematics educational practice can refer to mathematics itself, to semi-real or fictitious contexts, and real-life issues. On the other dimension, the learning environment can be within the exercise paradigm or landscapes of investigation. In retrospect, by examining projects developed in Skovsmose (1994); Alrø and Skovsmose (2002), Skovsmose advocates for opening new educational possibilities moving from the exercise paradigm towards landscapes of investigation, mainly on the grounds of allowing processes of interaction and communication to unfold. Moreover, if the purpose is to enable critique in the sense of reflecting upon the formatting power of mathematics in society, the movement in the other dimension ought to be towards real-life references. However, he warns that moving diagonally from 1 to 6 (see Table 2.1) implies moving from zones of comfort to zones of risks for both student and teachers. 'Uncertainties emerge. (...) However, a risk zone is a zone of possibilities" (Skovsmose, 2011, p. 48), and my project aims to explore new possibilities.

	Sequences of exercises	Landscapes of investigation
References to pure mathematics	1	2
References to a semi-reality	3	4
Real-life references	5	6

Table 2.1: Milieus of learning (Skovsmose, 2011, p. 40)

According to Artigue and Blomhøj (2013, p. 797) "inquiry-based pedagogy can be defined loosely as a way of teaching in which students are invited to work in ways similar to how mathematicians and scientists work". In their conceptualisation, they admit that this paradigm resonates with several well-established theoretical frameworks in mathematics education, including the dialogic-critical approach developed by Alrø and Skovsmose (2002). Blomhøj (2013) rounds the inquiry approach to three main points:

(a) there is something to be investigated, posed often as a guiding question as a point of departure,

(b) mathematical and pedagogical requirements are established for students' investigative work, and

(c) the students' results or reflections can provide a basis for building a relevant common knowledge.

An inquiry-based approach can provide structure to the learning environments, but "the central problem of an education based upon experience is to select the kind of present experiences that live fruitfully and creatively in subsequent experiences" (Dewey, 1938/2015). It is necessary to define criteria to select the exemplary investigations. Exemplarity, understood as the quality of an example, can be viewed from at least three perspectives: subjective exemplarity (based on Dewey's (1938/2015) philosophy of experiential learning), instrumental exemplarity (based on the definition of Wagenschein (1956)) and critical exemplarity (as interpreted by Negt (1975)). Respectively, an example acquires its quality depending on its resemblance to personal experience, its connection to disciplinary concepts and methods, and to the broader political and historical contexts. I develop these notions further in subsection 4.2.1, not to be accused of merely namedropping.

The general *problématique* in the thesis is focused on actual teaching-learning practice, so pragmatic considerations are also taken, including time limitations, previous knowledge and adherence to the curricular framework. These factors integrate crucial element highlighted by Skovsmose and Borba (2004) in their discussion of research methodological approaches compatible with critical mathematics education, referred to as *practical organisation*. Drawing from notions of Critical Theory as a research paradigm, they recognise that critique, transformation, advocacy, and activism are key concepts (Guba & Lincoln, 1994) to keep in mind. However, Skovsmose and Borba do not believe that the transformations needed in the classroom are decided *a priori* by the researcher. Considering this, an imagined or ideal learning situation is turned into an arranged or actual situation through a practical organisation in close collaboration of the teacher as a *subject* of research, instead of an *object* (Freire, 1970/1993). Moreover, the cycle is then updated by creating a new current situation, upon which a new imagined is proposed, and so forth. After all "the utopian aspirations of the critical paradigm may never be realised, but a more democratic society may materialise" (Scotland, 2012, p. 13).

These design principles, which I discuss further in Section 4.2, do not work as a checklist where all of them must be checked. One can anticipate some tensions between them, which, in light of the classroom experiences I will revise in Chapter 9. Consequently, the design principles play a role of guidance (Niss, 2007a) or framework for action (diSessa & Cobb, 2004; Bakker & van Eerde, 2015), as well as that of a local theory that can bridge problems with practice, in the sense that "a theory can help envision the inconveniences and puzzles that might result from the unfolding of new practices" (Silver & Herbst, 2007, p. 59).

2.3.3 Frameworks for analysis

A fundamental question in the thesis is whether a teaching practice can evoke the preoccupations of CME amongst students, in particular, that of the formatting power of mathematics. For this, it becomes necessary to delineate what does it mean to *critique* mathematics in classroom activity, i.e. to address **RQ4**.

Many versions of the meaning of critique can be found. For example, *checking* and *critiquing* are portrayed as cognitive skills on the revised Bloom's taxonomy (RBT), belonging to the broader category of *evaluating*. While checking is defined as "testing for internal inconsistencies or fallacies in an operation or act" (L. W. Anderson et al., 2001, p. 83), critiquing involves "judging a product or operation based on externally imposed criteria and standards" (p. 84). As an example, L. W. Anderson et al. (2001) draw the line between judging whether conclusions follow a scientist's method and whether the method is appropriate to solve the problem. In principle, this conception of critique is not compatible with the preoccupations of CME, as it does not account for the social consequences of the uses of a method.

A recent study by Radmehr and Drake (2019) makes connections between the RBT and influential theories of mathematics education, namely Skemp's theories of mathematical understanding, conceptual and procedural knowledge in mathematics, process-object theories (e.g. APOS), metacognition, the SOLO taxonomy, and Tall's three worlds of mathematics. The broader *evaluating* cell in the RBT can be considered as a metacognitive skill, "associated with checking the outcome of cognitive processing" (p. 911), no specific connections to *critiquing* are found.

Alrø and Skovsmose (2002) discuss that their version of a critical epistemology is rooted in aspects of Piaget's (1970) radical constructivism and Vygotsky's (1978) sociocultural approach. On the one hand, they share the notion of the learner as getting to know, seeing learning as a process, thus prefer using the term *knowing* instead of knowledge. On the other hand, they recognise learning as a collective or dia-logical (sic) endeavour, mediated by communicative resources. However, none of these inspirations takes on the task of critiquing mathematics. Alrø and Skovsmose (2002, p. 294) characterise a critical epistemology as "a theory of developing or constructing knowledge, where critique of what is learned is seen as part of the learning process." In that regard, Skovsmose (1994, Chapter 6) lays out six entry points on reflective knowing to offer a basis for its operationalisation, formulated as archetypal questions:

1. Have we done the calculation right?
2. Have we used an appropriate algorithm?
3. Are the results reliable for the purpose we have in mind?
4. Is is appropriate to use a formal technique at all? Do we, in fact, need mathematics?
5. How does the application of an algorithm affect our conception of a part of the world?
6. Could have we reflected upon the use of mathematics in another way?

These entry points differ according to the scope of reflections. In Skovsmose (1998), he refers to the first and second entry points as mathematical-oriented reflections, the third and fourth as model-oriented reflections, and the fifth and sixth as context-oriented reflections. The idea behind reflective knowing is to work as an extension of Dewey's notion of reflective inquiry to include a critique of science and mathematics itself. In this thesis, these entry points help me to analyse the scope of student's reflections during the development of classroom experiences. The main focus is on the fifth entry point, namely on the formatting power of mathematics as a fundamental preoccupation of CME.

Domain-specific: Probability and statistics

The *problématique* of the thesis is specifically focused on the critical justification of probability and statistics in upper secondary school. Therefore, an additional domain-specific layer to the analysis is given for each of the classroom experiences, to address **RQ5**.

In Chapters 5 and 6 follow the commentary by Batanero (2002) in regard to broad models for statistical (eventually probabilistic) literacies. She claims that those may be useful for addressing curricular priorities in policy-making and as general building blocks. However, models at a micro-level may be needed for research in specific learning environments.

In Chapter 5, I make use of a local framework directly connected to the inquiry at stake. The inquiry refers to a real judicial case where probabilistic notions were used to convey evidence (People v. Collins, 1968), analysed by Rosenthal (2005, 2015). Simply put, a suspect couple is detained on the grounds of matching the description given by a witness. The probability of a couple matching the features is computed. As a result, its low value is used as evidence for a conviction. Rosenthal (2015) defines three issues to question the use of that probability. First, the question of "when to multiply?" addresses the rule of multiplications of probabilities. In the case described, features need to be independent of each other. The character of independence is a significant aspect in frameworks of probabilistic thinking and literacy (Borovcnik, 2017). Second, the prosecutor's fallacy (Thompson & Schumann, 1987) is the misinterpretation of what is actually computed, namely $\mathbb{P}(\text{evidence}|\text{innocence})$, as if it were the likelihood of the suspect's innocence based on the collected evidence, i.e. $\mathbb{P}(\text{innocence}|\text{evidence})$. This inequality is exemplary of the larger issue of asymmetry of conditional probabilities (Borovcnik, 2006). Third, the "out of how many?" principle refers to the consideration of multiple tests from a frequentist perspective (Batanero & Díaz, 2007a). The probability of matching the suspects' features might be small, but finding more than one such couple in a large population could still be not so rare.

In Chapter 6, I also make use of a local framework for analysing graphical representations with a political context (Table 6.4). In fact, this framework is based on students observations, which we organise with the teacher during the classroom experience to form this proto-framework for analysis. There is, of course, a body of research into the interpretation of graphical representations of data (e.g. Curcio, 1987; Friel, Curcio, & Bright, 2001). However, they refer mostly to cognitive skills without much attention to critical issues or the relationship between the learner with context (Monteiro & Ainley, 2007). Niss (2007a, p. 105) describes this type of use of theories (I am abusing the term *theory*) to "focus on *organising a set of specific observations and interpretations* of singular but related phenomena into a coherent whole". Niss argues that this type of bottom-up approach is the case of grounded theory (or meta-theory), an approach to build and develop theory from empirical observations in a systematic way, registering memos, coding in sequential phases and applying criteria to achieve theoretical saturation (Teppo, 2015). I do not claim this to be the ultimate aim, nor the actual carrying out it implies. The proto-framework intends to delineate a set of considerations that matter when reading and constructing graphical representations in the context of the specific task.

In Chapters 7 and 8, I take a more top-down approach for the analysis by trying out recently proposed frameworks of probabilistic and statistical literacy, namely the ones proposed by Borovcnik (2017) and Weiland (2017). Here, risk-probabilistic and critical statistical literacies play the role of "overarching frameworks from which (parts or aspects of) the teaching and learning of mathematics can be viewed and approached" (Niss, 2007a, p. 104). Part of this shift from the previous Chapters is due to the exploratory character of the project. I want to test if and how domain-specific literacy frameworks facilitate a more direct connection between probability and statistics, and the notions of reflective knowing of Skovsmose (1992, 1994).

The first use of the term "probability literacy" that I am aware of is the one proposed by Gal (2005), as a construct composed by knowledge (big ideas, figuring probabilities, language, context) and dispositional elements (critical stance, beliefs and attitudes, personal sentiments). The contextual element refers to being aware of the "omnipresence" of chance and probability in the world, and the critical questions refer to the context, source, process, meaning and interpretation of messages, in a right-wrong divide. Scholars have moved forward in the development of probabilistic *thinking* instead, opening to broader perspectives, including humanistic, psychological, stochastic and educational (Chernoff & Sriraman, 2014b). One of the most comprehensive frameworks in probabilistic thinking is that of Jones, Langrall, Thornton, and Mogill (1997), validated in further empirical research in Russia and Lesotho (Mooney, Langrall, & Hertel, 2014). However, the judgement of probabilities is only a part of exploring notions of risk, as we do in Chapter 7.

Borovcnik and Kapadia (2014) provide an extensive historical and philosophical revision of the development of probability (see also Section 3.1 of this document), situating modern philosophical perspectives as focusing on two main aspects: the connection between probability and physics, and notions of risk. Their revision reveals that "risk has always played a prominent role in the development of the (probabilistic) concepts" (p. 29), to the lengths of that risk and decision making define the logic of probability (Borovcnik, 2015). Borovcnik (2017) unfolds a further discussion of probabilistic thinking and literacy in the context of risk, by critiquing Gal's (2005) model and to "corroborate how interrelated probability and risk are and how literacy in probability is entrenched by literacy in risk" (Borovcnik, 2017, p. 1497) in several contextualised examples of decision making. Moreover, he reorganises previously developed ideas (Borovcnik, 2011; Batanero & Borovcnik, 2016) to "perceive probabilistic literacy as the ability to use relevant concepts and methods in everyday contexts and problems" (Borovcnik, 2017, p. 1500).

On one dimension, there are four *abilities* of probabilistic thinking relevant in the context of risk, related to address competing intuitions (Borovcnik, 2011), namely:

- The ability to balance between *psychological* and *formal* elements
- The understanding that direct criteria for success are missing
- The ability to separate between *randomness* and *causality*
- The ability to separate *reflecting* on a problem and making a *decision*

On the other dimension, probability literacy in risk implies making use of secondary intuitions, based on the *themes* described in Batanero and Borovcnik (2016), namely:

- Theoretical character of probability and independence:
  - Meanings of probability: subjective (SJT), *a priori* (APT), and frequentist (FQT) theories of probability (Borovcnik & Kapadia, 2014)
  - The assumption of independence
  - The problem with small probabilities
- Conditional probability:
  - Dependence on prior judgements and the Bayes' formula
  - Asymmetry of conditional probabilities
- Concepts building on probabilistic evidence, such as probabilistic dependence as conceptualised by the correlation coefficient or relative odds.

The choice for Borovcnik's (2017) version of probabilistic literacy is on the grounds of being the most recent and risk-oriented framework.

Finally, in Chapter 8, I looked for a framework that could account for the critical stance in statistical literacy. Given by Gal, and still quoted (e.g. Garfield & Ben-Zvi, 2008; Sharma, 2017; Ben-Zvi, Makar, & Garfield, 2018), statistical literacy can be defined as

> people's ability to *interpret and critically evaluate* statistical information, data-related argument... to *discuss or communicate* their reactions to... statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions.
>
> (Gal, 2002, pp. 2–3, emphasis in original)

Although diverse definitions and characterisations have been given to statistical literacy, its components usually relate to technical knowledge, literacy skills, and critical attitudes. Gal's seminal model for statistical literacy (Gal, 2002, 2004) divides components into knowledge and dispositional elements. Within the knowledge components, he includes critical or worry-questions, mostly related to sources of data, validity and reliability of methods, and alternative interpretations. As a dispositional element, Gal proposes a critical stance as a propensity of adults to ask critical questions, leaving calls for action out of a framework dedicated to consumers of statistical messages.

Watson and Callingham (2003) provide a hierarchical construct based on the SOLO taxonomy (Biggs & Collis, 1982), where, in the two highest levels, the individual's engagement with context is, respectively, critical and critical-mathematical. Here, critical refers to a questioning attitude with context, appreciating chance and variation, and language vigilance. Sharma, Doyle, Shandil, and Talakia'atu (2011) define and try out a hierarchical model to inform the design of classroom activities in four stages. The higher two stages, early and advanced critical, provoke questioning features of data and context in diverse perspectives. In both models, the higher level is distinguished by the display of mathematical and statistical skills to inform critique. Sharma (2017) offers a thorough literature review of frameworks of statistical literacy, highlighting that they are, at most, dedicated to reading contexts, i.e. positioning the citizen as a reader of statistical messages. The tasks she makes use of in Sharma et al. (2011), for example, are all paper-and-pencil tasks where the statistical information, interpretations and representations are already presented instead of produced by the students.

As we shall see, one lesson from Chapter 6 is that there are differences between reading and writing graphical representations in terms of reflective knowing. On that note, Weiland (2017) provides a problematisation of statistical literacy discussing three aspects: the use of the term critical; the dichotomy of consumer (in reading contexts, see Gal (2002, 2004)) versus producer (in inquiry contexts, see Wild and Pfannkuch (1999)) of statistical messages; and the relevance of context. He draws in Freire's more emancipatory signification of literacy. In that sense, Weiland is explicitly critical of Gal (2002, 2004), stating that "the notion of enabling action needs to be pushed farther as well, to not only consider enabling action as a result of reading statistics, but to also enabling action by writing the world through statistics" (p. 39). Furthermore, he proposes a framework for critical statistical literacy, as an intersection of statistical literacy, with Gutstein's (2006) notions of reading and writing the world with mathematics. The framework is summarised in Table 2.2.

Reading the world with statistics	Writing the world with statistics
Making sense of language and statistical symbols systems and critiquing statistical information and data-based arguments encountered in diverse contexts to gain an awareness of the systemic structures at play in society.	Using statistical investigations to communicate statistical information and arguments in an effort to *destabilise and reshape* structures of injustice for a more just society.
Identifying and *interrogating* social structures which shape and are reinforced by data-based arguments.	Using statistical investigations to *alleviate* and resolve sociopolitical issues of injustice
Understanding one's social location, *subjectivity*, political context and having a socio-historical and political knowledge of self and understanding how it influences one's interpretation of information.	*Communicating* one's social location, subjectivity, and political context to others and how it shapes one's meaning-making of the world when reporting results of a statistical investigation.
Evaluating the source, collection and reporting of statistical information and how they are influenced by the author's social position, and sociopolitical and historical lens.	*Negotiating* societal dialectical tensions when formulating statistical questions, data collection and analysis methods and highlighting such tensions in the results of a statistical investigation.

Table 2.2: Critical statistical literacy framework (Weiland, 2017, p. 41)

Mathematics education scholarship builds and uses multiple theories and frameworks at different levels of reach and often combined. Probability and statistics education is no exception (Chernoff & Sriraman, 2014a; Nilsson et al., 2018). In this chapter, I have discussed my overall conceptual framework to address the five research questions that pertain to the general *problématique*.

Critical discourse analysis serves the purpose of making research questions **RQ1** and **RQ2** approachable, defining a coherent analytical strategy. The three-dimensional model proposed by Fairclough (1992, 2003) positions texts (the actual) as observable cases of broader discursive practices (meaning-making) that shape social structures (the possible). From the point of departure of attempting to make sense of selected curricular texts that refer to *decision-making*, CDA provides a research methodology for an empirical study (Niss, 2007a).

In order to keep a resonance with the philosophy of critical mathematics education, I draw on inquiry-based mathematics education as a framework for action, since "from its origins, an essential aim of IBE (inquiry-based education) has been to promote values of emancipation and democracy" (Artigue & Blomhøj, 2013, p. 808). It is grounded in the educational philosophy of experiential learning developed by Dewey (1938/2015), and a fundamental task is to define adequate experiences for the learning purposes. This challenge leads to the principle of exemplarity, which can be seen from different philosophical perspectives. In this thesis, I consider the subjective perspective of Dewey, the instrumental perspective of Wagenschein (1956) and the critical perspective of Negt (1975). Moreover, in order to adhere the designs to actual teaching practices, the principle of pragmatism or pragmatic organisation can lead to turning an imagined learning situation into an arranged pedagogical situation (Skovsmose & Borba, 2004). These principles do not only serve as guidance for action or behaviour (Niss, 2007a), but will also illustrate the tensions and puzzles after the unfolding of the teaching practice (Silver & Herbst, 2007), thus addressing **RQ3**.

The philosophy of critical mathematics education (Skovsmose, 1994), with the further development of a critical epistemology of mathematics (Alrø & Skovsmose, 2002), serves as an overarching framework from which the classroom experiences are designed and viewed. In particular, Skovsmose's (1992, 1994) entry points on reflective knowing are the basis for the analysis, allowing to interpret whether and how students reflect upon the formatting power of mathematics. Moreover, the dialogical features of the inquiry-cooperation model developed in Alrø and Skovsmose (2002) provide the means for describing episodes in the classroom experiences.

In Chapter 5, I use a specific framework to the judicial case studied (Rosenthal, 2015). In Chapter 6, we take the opportunity to develop a local proto-framework for analysing graphical representations, grounded in the classroom experience. In Chapter 7, I use a two-dimensional framework for probabilistic literacy in the context of risk, proposed by Borovcnik (2017). Finally, in Chapter 8, I use the framework of critical statistical literacy presented by Weiland (2017).

I began this chapter by referring to the point made by Goos (2018) about the multiplicity and entanglement of theoretical underpinnings and frameworks in mathematics education research. The conceptual framework of this thesis is no exception.

Chapter 3

The justification problem

The first part of the research project is that of exploring the diversity of reasons and purposes for teaching mathematics and, in particular, probability and statistics in high school. In the first *International Handbook of Mathematics Education*, Niss (1996) begins his address of the justification problem with the following quote:

> A discussion of mathematical education, and of ways and means of enhancing its value, must be approached first of all on the basis of a precise and comprehensive formulation of the valid aims and purposes of such education. (...) Such aims and purposes of the teaching of mathematics, moreover, must be sought in the nature of the subject, the role it plays in practical, intellectual, and spiritual life of the world, and in the interests and capacities of the students.
>
> (National Committee on Mathematical Requirements, 1923, cited in Bidwell & Clason, 1970, p. 390)

The point is that, if we intend to capture the big picture of research in some branch of mathematics education, we need to start by addressing the questions *why* and *for what* we are teaching and learning it. Such purposes, as established in the quote above, ought to account for the nature of the subject matter itself, the role it plays in the world and the students' position.

In general, (real) reasons refer to the "driving force, typically of a general nature, which in actual fact has motivated and given rise to the existence (i.e. the origination or the continuation) of mathematics teaching withing that segment" (Niss, 1996, p .12). A reason addresses the question of *why* to teach. The parenthetical suffix (real) added by Niss makes the point that reasons are not necessarily *good*, as they are often complex, vague, fuzzy and inertial. That is the case of many historical reasons, which are entangled within broader sociopolitical processes, geopolitical included. For example, the influence of 19th century European intellectuals in the (Latin) American independentist movements, can be well be raised as a reason to include statistics as a core subject in Chilean universities, as illustrated in Andrés Bello's inaugural speech:

> The University will examine the results of Chilean statistics, it will contribute to shaping it, and will read in its numerals the expression of our material interests. Because in this, as in other subjects, the University' programme is entirely Chilean: if it borrows from Europe the deductions of science, it is to apply them to Chile.
>
> (A. Bello, 1843, p. 147, my translation)

Similar reasons for the inclusion of teaching standards and practices in Latin America – for mathematics in general (de Carvalho, 2014) and stochastics in particular (Cuevas, 2012) – during the second half of the 20th century can be traced back to the *new math movement*, as Latin America was one of the battlegrounds of the Cold War.

When reasons are activated or expressed, then they turn into justifications or arguments. For these arguments to make sense, they are accompanied by – more or less explicit, more or less substantiated – claims for a reason to be solid and factual. Following up on the previous example, it was not until 1955 that statistics began to be taught in the University of Chile, following the argument that statistics "constitutes an instrument of recognised efficacy in the analysis and resolution of economic and practical problems, as well as in the formulation and establishment of plans for economic development of the countries" (Cansado, 1955, p. 121). This justification for teaching statistics in Chilean academia comprises not only that other countries do in fact develop their economic plans employing it, but also the explicit claim that it is recognisably effective for those purposes.

A distinct concept is that of *goals* for mathematics teaching, encompassing terms such as ends, aims, objectives and purposes (Niss, 1996). Goals address the question of *for what* to teach. In a way, this is the discussion giving birth to frameworks of probabilistic and statistical literacies and their connections to school education: "educators have to distinguish between teaching more statistics (or teaching it better) and teaching statistics *for a different (or additional) purpose*" (Gal, 2002, p. 21, emphasis in original). Goals express the final outcome one hopes to achieve by teaching a subject. Their most explicit form take place in official curricular frameworks. School study programmes organise the purposes and expectations for students to achieve by participating in teaching activities of each specific subject. These are often enlisted as explicit learning goals.

This chapter deals with the justification problem of probability and statistics in high school-oriented to notions of decision making. In Section 3.1 I make a brief historical account on the role of decision making as a driving force in the development of probability and statistics as scholarly knowledge, thus accounting for the nature of the subject matter. In Section 3.2 I address the psychological agenda of heuristics and biases in decision making under uncertainty and risk, and its relation to probability and statistics education. Here I address the justification problem from an individual's position, as portrayed in the literature. In Section 3.3 I dig into what is elicited in curricular frameworks about decision making and the teaching of probability and statistics in high school. It serves two purposes. The first is to give a proof of existence to the claim that decision making is part of the justification, beyond the school sphere, and it is embedded in curricular documents. The second purpose is to dig deeper into how decision making conveys legitimacy to the inclusion of probability and statistics in curricula. In Section 3.4, I come back to Niss's frame and choose a particular type of justification to keep working on this thesis as a teaching purpose, namely the critical justification.

3.1 The history of probability to mathematise decisions

The history of probability as a scientific and mathematical field of knowledge is entangled with the history of decision making when uncertain outcomes are to be evaluated. The idea of making decisions utilising mathematical models seems trivial today, as the application of yet another mathematical theory to inform decisions can be taken for granted. However, the mathematical formalisation of probability theory did not take place until the beginning of the 20th century by Andrei Kolmogorov (1933/1956), while the history of decision problems leading to notions of probability and statistics go a long way. I offer a brief trail through the history probability and statistics, highlighting its milestones and their relation to decision making. As primary guidance, I draw on the extensive work by Hacking (1975, 1990), the historical and philosophical perspective on probability by Borovcnik and Kapadia (2014), and the summary on the history of statistics by Wild et al. (2018).

From ancient history, games are an indisputable origin of notions of chance. Scholars believe that the first dice were constructed around 3500 BC with an astragalus (a heel bone from sheep) and there is evidence of Babylonian pottery cubes from around 3000 BC (David, 1962). From early times to the Middle Ages, chance had the role of mediating between people in conflict and to link divinity with men in religious rituals, with the use of astragali or drawing beans form urns. These objects had the function to "decide fairly or to explore god's will for a decision" (Borovcnik & Kapadia, 2014, p. 9). Chance was the manifestation of divinity's saying. The early origins of chance are connected to decision making as being it the decision-maker.

The roots of mathematical probability did not come up until the 17th century, being Blaise Pascal and Pierre de Fermat's epistolary exchange a turning point of interest. They and others in the Roannez circle (Hacking, 1975, Chapter 7), confronted and solved problems involving games of chance, whereby equiprobability seemed not to be applicable, as much as they were based on rolling dice and tossing coins. I refer namely to de Méré's problem and the division of stakes. De Méré's problem poses a choice to be made between two – apparently equally favourable – games of dice. The problem of the division of stakes, in general, is that of a chance game between two players with a limited number of rounds and defining the fair share between them if the game is interrupted at a certain point. This puzzle is an early version of the problem of pricing, i.e. of assigning monetary values in the situation on whose fairness both parties agree. At this point, probability calculations were not yet involved: "Pascal, Fermat, and Huygens were concerned with a problem of equity, not a problem of probability. They were pricing gambles, not evaluating evidence or argument." (Shafer, 1996, p. 16). Their contribution is figuring out that the combinatorial multiplicity of outcomes or possibilities can be favourable to one player. De Méré's problem consists of deciding between two games. In the division of stakes, as the pricing problem it is, the question is at which price one decides to play the game. It is not a fair God who decides anymore. It is the player who decides to participate in a game of chance.

The division of stakes problem illustrates how the decision process does not only concern chances, but also the stakes involved. Not only likelihoods are to be determined, but their impacts as well. This addition raises the next milestone in the development of probability: expectation (Hacking, 1975, Chapter 11). Christiaan Huygens, who authored the first printed book of probability, attempted a formulation for the worth of a situation to be a weighted average between possible outcomes and their respective chances (understood as *a priori* frequencies or multiplicities). Explicitly, Huygens also included probability in a "theory" on decisions, though "probability was not yet a number, but a collection of arguments (pro and con) in order to weigh these arguments" (Borovcnik & Kapadia, 2014, p. 14). On a parallel track, Huygens was a statistician and worked with mortality tables for the pricing of annuities, as did John Graunt in England. These developments set the course for a frequentist meaning of probability, more established by Jakob Bernoulli's version of the law of large numbers. This period contributes to mathematise the price of a gamble in decision problems. Moreover, Huygens and Graunt's application of the frequentist approach to collect, organise and make decisions upon data at a larger scale, represents the gestation of statistics, the arithmetic of the State.

The practice of statistics as a comprehensive collection and description of data in the form of censuses is as old as paper, but the birth of modern statistics can be associated to the consolidation of European nation-states during the 18th century. The term *Statistik* was coined as an alternative to *Staatswissenschaft* (Achenwall, 1749) and translated to English as political science or political arithmetic. According to Donnelly (1998, p. 231), "the political arithmeticians made their enquiries with an explicit view to considerations of the state power". In that sense, Graunt can be declared the first statistician (Wild et al., 2018). In line with the ideals of modernity, the driving force was the idea that state policies ought to be the determined by making sense of data and not by old-fashioned sources of authority such as the church or nobility (Porter, 1986). In consequence, the advancements of the notions of empirical probabilities would inform decisions beyond personal gain, but at a broader State-level.

Borovcnik and Kapadia (2014) define two obstacles for further development of probability. Firstly, the so-called St Petersburg Paradox, a game of chance whose expected value diverges, leading Daniel Bernoulli to distinguish between mathematical and moral expectations (Mateos-Aparicio, 2002, p. 9), thus implanting the modern idea of utility developed further during the 19th-century (Mateos-Aparicio, 2004). The second turning point was materialised by Bayes' formula and the notion of inverse probability, i.e. that one can obtain the probabilities of the causes by which an observed event could have been produced. Bayes' method, based on probabilities interpreted as relative frequencies, "facilitates the correction of subjective probabilities, a basic element in decision theory" (Mateos-Aparicio, 2002, p. 13).

The classical formalisation of probability theory can be attributed to Pierre-Simon de Laplace, who, among many contributions, engages in the problem of statistical inference over a population. Further development in continuous distributions by Abraham de Moivre and Laplace's central limit theorem paved the way for Carl Friederich Gauss' application of the normal distribution to the least-squares method for observational errors. Though the applications in this period began to be closer to astronomy and other branches of physics than to chance games, the least-squares method is by itself an optimisation problem. The method consists in minimising the sum of square errors, and both "Laplace's approach and his (Gauss') are particular cases of a more general loss function" (Ríos, 1998, p. 19). The minimisation of loss and maximisation of expected utilities are at the core of developments in decision theory during the 20th-century (Mateos-Aparicio, 2004).

By the beginning of the 19th century, three factors colluded. First, large amounts of data were accumulated, which Hacking calls the "avalanche of printed numbers" (1990, p. 2), and it led statisticians to see new patterns. Second, the conceptual framework shifted from the political (sovereignty and power) to societal (autonomy and population). Third, the advancements in probability theory and statistical inference began to be interpreted as law-like regularities, representing divine providence or natural design, even for moral facts, such as homicides and suicides (Hacking, 1990, Chapter 8). These factors produced the shift from political arithmetic to social statistics, i.e. the original social science (Donnelly, 1998). The implication was that *statists* broadened the scope of investigations to trade, industrial progress, labour, poverty, education, sanitation and crime (Porter, 1986). This diversity of areas of expertise can be illustrated in the early statistics societies (London Statistics Society and American Statistical Society). Nowadays, statistics can be applied to an infinite variety of contexts. In discussing why statistics is more important than ever, Wild and colleagues argue that "statistical ideas and methods provide many tools for making decisions, in life, especially decisions that involve trade-offs" (Wild et al., 2018, p. 26) and highlight the relevance of expected values and hypothesis testing as crucial ideas for decisions makers to know.

Overall, the development of probability and statistics cannot be separated from the notion of decisions. On the one hand, decision problems led to establishing early probabilistic ideas and calculations. On the other hand, developments of probability and statistics are at the core of decisions both in theory and practice. A more contemporary way of defining this intrinsic relationship is through the notion of risk. This framing is how Borovcnik and Kapadia (2014) conclude their historical and philosophical address:

> As may be seen from the historical development, risk has always played a prominent role in the development of the concepts. This is not restricted to games of chance but extends to more general situations under uncertainty. (...) In some situations, some outcomes have a small impact or may have a benefit. The decision situation involves several options, all with different outcomes related to the impact (loss or win). Which of the decisions is best?
>
> (Borovcnik & Kapadia, 2014, p. 29)

Moreover, who is to say? In today's world risk society, the explosive diversity of fields of knowledge and internationalisation of responsibilities make utmost relevant for every citizen to participate: "how one acts in this situation is no longer something that can be decided by experts" (Beck, 2000, p. 217). A justification for probability and statistics in school mathematics aligned with the cultural maintenance of the field, ought to be connected to risk and decision making. That is, a justification rooted in the nature of the subject matter should be rooted in "risk and decision making: the *logic* of probability" (Borovcnik, 2015).

3.2 The elephant in the room: Heuristics and biases

In the process of publishing a journal article (Elicer, 2020), one of the reviewers – certainly the most acquainted to probability and statistics education research – insisted on the necessity of making an up-front mention to the psychological tradition of decision making. Though this chapter intends to discover how decision making is invoked in the curricular texts themselves, instead of me providing preconceived justifications, when it comes to discussing the relation between decision making and school probability and statistics, it is unavoidable to bring up what Batanero, Chernoff, Engel, Lee, and Sánchez (2016) call the *heuristics and biases programme*. Let me address the elephant in the room.

As a root for explaining probabilistic intuitions and difficulties, this educational research programme is inspired by a cognitive and behavioural research agenda in psychology referred mostly to the work of Daniel Kahneman and Amos Tversky. As a seminal example, Kahneman and Tversky (1979) define a prospect theory of decision under risk, by acknowledging and including four intuitive effects (certainty, reflection, probabilistic insurance, and isolation) that violate the axioms of ever-popular utility theory (Friedman & Savage, 1948). That is, they construct a new model of decision opposed to the assumption that people decide by simply maximising expected utility. Overall, by focusing on decision making under risk, researchers identify types of intuitive conceptions (heuristics) that lead to systematic errors (biases), in contrast to what would result of making unbiased, rational choices.

In his widely known dissemination book *Thinking, fast and slow*, Kahneman (2011) summarises his and Tversky's work, defining a two-system mind theory. System 1 corresponds to the fast, intuitive, heuristic thinking, and System 2 is the slow, deliberate and rational thinking. According to this theory, these two systems are differentiated on the cognitive load involved, heuristics playing the role of mind shortcuts: "when faced with a difficult question, we often answer an easier one instead" (Kahneman, 2011, p. 12). A persistent element in this agenda is the focus on individual and economic-inclined types of decision, that presumes a rational, correct alternative.

These findings provide a rich palette of resources to identify probabilistic misconceptions and its consequences (e.g. Kahneman, Slovic, & Tversky, 1982; Borovcnik, 2017). This work has been given high regard by the probability and statistics education community. There are several works identifying students' probabilistic misconceptions (e.g. Serrano, Batanero, Ortíz, & Cañizares, 1998; Batanero & Sanchez, 2005) and errors in statistical reasoning (e.g. Garfield, 2003). However, the beneficial role of education is not trivial, and studies have shown how the incidence of some errors decline throughout formal education, others persist, and some biases even grow with age, and educational level (e.g. Fischbein & Gazit, 1984; Fischbein & Schnarch, 1997). As Garfield and Ben-Zvi sum up, these studies show that "inappropriate reasoning about [probabilistic and] statistical ideas is widespread and persistent, similar at all age levels (even among some experienced researchers), and quite difficult to change" (2007, p. 374).

Other researchers have a different take on intuitive decision making. Kruglanski and Gigerenzer (2011) revise different decision strategies and realise that both intuitive and deliberate judgements can be modelled as following the same rules. That is, They construct a 2-step process for rule selection, the first being constrained by the task and the person's memory, and the second determined by the individual's process capacity and perceived ecological rationality. Ecological rationality refers to "the environmental structures in which a given heuristic is successful – that is, the match between mind and environment" (Gigerenzer, 2008, p. 23). Instead of focusing on heuristics as sources of error, their approach is to understand under which environmental conditions heuristics do work.

The latter approach has also influenced the probability and statistics education field, informing research addressing the role of context in decision making under risk. This body of research has developed an understanding of so-called errors as a correct answer to a different question that makes sense in the proper context. For example, Pratt and colleagues explore the complexity of deciding whether Deborah (a fictional subject) should have an operation that could cure her medical condition but, at the same time, carries underlying risks (Pratt et al., 2011; Levinson, Kent, Pratt, Kapadia, & Yogui, 2012). By exposing a group of teachers to the dilemma, they reveal how context affects the weighing of likelihoods and impacts. In a sequence of dice games with 7th-grade students, Nilsson (2007) uses a contextualisation perspective to consider how students' understanding is reorganised as the task evolves and their previous strategy becomes obsolete. This strategy is illustrated not only by the decisions they make but also in their dialogues. Moreover, Prediger (2008) shows how a student is aware of the (apparent) gap between her "probability" thinking and "normal" thinking. She proposes that making different conceptions and strategies explicit, challenging them with experiments and becoming aware of underlying differences between them, should be part of adequate learning environments. On the grounds of research experiences about the complexity and scaffolding of decision-making tasks under uncertainty, Pfannkuch (2018) has identified emerging curricular approaches to be addressed in future research. One proposition is to get "*more insight into fostering statistical argumentation* including learning how to make evidence-based claims in data-rich environments and critically evaluating data-based arguments in diverse media from a statistical literacy perspective" (Pfannkuch, 2018, p. 407, emphasis in original).

The role of heuristics and biases is a significant and ongoing contribution to the understanding of decision making under uncertainty and risk. Identifying these pitfalls can explain some of the difficulties that students encounter when dealing with probability and statistics learning situations that involve making choices and taking actions. However, early work is focused on economic-oriented individual decisions and developed under laboratory conditions, while, as we will see later, the justification for the inclusion of probability and statistics in upper secondary school has a broader scope. Educational studies have illustrated how persistent these misconceptions are, despite growing up and following formal education. Moreover, when put into real-life contexts, one cannot rule out the rationality of a decision that may be ecologically rational. I do not disregard this research programme at all. I could not. In fact, it was at the core of the genesis of my research project and the core motivation to follow PhD studies. I am addressing the elephant in the room that I invited to come in in the first place. However, as we will see in the next section, the psychological agenda is just a part of the justifications given for teaching and learning probability and statistics in school.

3.3 Curricular frameworks as instances of legitimation discursive strategies

In order to keep walking through the general *problématique* towards an actual teaching practice, it is of particular interest to explore whether mathematics curricula reflect critical ambitions within a decision orientation in their inclusion of probability and statistics. Mentioning and connecting these notions can provide a thorough justification for the learning of data and chance in school for every future citizen.

In this section, I address the justification problem by investigating the role of decision making in current and upcoming probability and statistics official curricula, through a version of

Critical Discourse Analysis (CDA). In order to narrow down and have a sense of what is the aim regarding students' future participation as citizens, I make an exemplary analysis that focuses on upper secondary school mathematics curriculum. As a Latin American researcher in a Nordic environment, I take Chile and Denmark – my home and host countries, respectively – as exemplary cases. Accordingly, the work reported in Elicer (2019) focuses the analysis to the Chilean curricular framework, in consonance with the empirical work taking place in that educational environment. In Elicer (2020) I expand and delineate a comparative study between both countries. As a disclaimer, for the sake of the thesis' self-containment, I make use of those papers in what follows. They are partially paraphrased or reproduced, with the editors' permission.

The research questions addressed in this section are:

RQ1 What is the role given to notions of decision making in the Chilean and Danish upper secondary school probability and statistics curricula?

RQ2 Which similarities and differences can be drawn from these two cases?

These enquiries contribute to the field of probability and statistics education research by providing an exploration of how curricula reflect justifications that appeal to greater ideas of decision making, democracy and citizenship, thus moving towards a more transparent link between school mathematics and its declared goals. Interpreting roles and spotting changes are broad intentions, so in the following section I address the conceptual framework and methods concerning a version of Critical Discourse Analysis, in order to reinterpret the research questions in a more precise and operational way.

3.3.1 Analytical strategy

Inspired by Fairclough, I am describing the methodology as an altogether analytical strategy, since "we cannot so sharply separate theory and method" (2010, p. 234) while constructing the object of research. Starting with theoretical generalities, I then provide the necessary conceptual definitions required by the methods for selection and analysis.

Key interrelated concepts are social structures, practices and events, where "social structures define what is possible, social events constitute what is actual, and the relationship between potential and actual is mediated by social practices" (Fairclough, 2003, p. 223). In this frame language is the social structure. Among its infinite possibilities, choices are made to produce texts as part of social events, mediated by discursive practices, where discourse is understood as semiosis, i.e. the process of meaning-making.

In line with this theoretical framework, my analytical strategy consists of navigating backwards the language-discourse-text hierarchy. I first select texts from upper secondary school Chilean and Danish official curricula that refer to decision making in probability and statistics. I then extract elements of textual analysis, which finally illustrate broader discursive practices. The textual analysis describes what is in the texts, and the discursive aspect addresses how these elements give meaning. The upper level of social structures would involve a discussion on all there is possible to write and mean about decision making in the curriculum, through a comprehensive global-scale study, which is beyond the realm of this section.

Data collection

Chilean secondary education is defined from grades 7 to 12. These are called 7th-basic and 8th-basic, and from 1st-middle (9th-grade) to 4th-middle (12th-grade)[1]. All of these grades have their own study programmes, but more general curricular frameworks are available for 7th-basic to 2nd-middle, and the one for 3rd and 4th-middle is under construction at the time the analysis took place[2]. The text selection for Chile takes into account 3rd and 4th-middle curricula, and the upcoming curricular framework aforementioned.

For current versions of the curriculum, I will only take common plan programmes, i.e. those directed to every student in the school system. Excluded programmes are those differentiated mathematics studies defined for scientific education. Therefore, the texts are extracted from grades 3rd-middle and 4th-middle study programmes. These documents also contain didactical guidelines for teachers.

With regard to Denmark, upper secondary (*gymnasium*) education can be taken after ten years of mandatory basic education (*grundskole*). There exist four main gymnasium programmes, namely STX, HTX, HHF, and HF, and three possible levels of mathematics within them: A, B and C. In order to narrow down to one mainstream version of mathematics "for all", I select texts from study plans and guidelines for B-level mathematics. An additional document not authored but endorsed and published by the Danish Ministry of Education is included as a particular set of guidelines with respect to statistics.

All sources of texts are summarised in Table 3.1. For each source document, are given: a label, year of publication or last update, curriculum it belongs to, and citation on the reference list.

Document	Label	Year	Curriculum	Reference
Mathematics Study Programme, Third Middle	CL315	2015	2009	MINEDUC (2015b)
Mathematics Study Programme, Fourth Middle	CL415	2015	2009	MINEDUC (2015a)
Curricular proposal for 3rd- and 4th-middle: Public consultation document	CL3417	2017	2019	MINEDUC (2017)
Mathematics B Study Plan, HTX	DKB17	2017	2017	UVM (2017)
Mathematics A/B - HTX: Guidelines	DKAB18	2018	2017	UVM (2018b)
Mathematics B/C - HF: Guidelines	DKBC18	2018	2017	UVM (2018c)
Mathematics A/B/C - STX: Guidelines	DKABC18	2018	2017	UVM (2018a)
Statistics in high school	DKSG17	2017	2017	Ekstrøm, Ernst, and Brockhoff (2017)

Table 3.1: Sources of selected texts

I look for sentences making references to ‘decision’ and ‘choice’; respectively, *decisión* and *elección* in Spanish, and *beslutning* and *valg* in Danish, including conjugations of the verbs ‘to decide’ and ‘to choose’. These sentences compose the texts to be analysed.

Chilean study programmes are divided into units (numbers, algebra, geometry, and data and chance). I focus the scope of the search on the units of data and chance. The same selection is applied to the upcoming curricular proposal for these last two grades (CL3417), within the mathematics subject. It still does not have the form of study programmes, but it can be considered to be of the same genre, for it means to become the next version of such. It is a general proposal based on a process that involved national experts, international experiences and consultation with civil society participants. The document is published on MINEDUC’s website.

Danish study plans and guidelines are not divided into grades, and probability and statistics appear among learning goals instead of main educational unit headlines, making the scope of the search in the full documents broader.

Not all appearances – close to 100 – are included in the selection, since my purpose is to investigate links between decision making and subject matter probability and statistics. Therefore, mentions of pedagogical choices in content and methods, or students’ decisions for a particular exam are discarded. At the same time, as seen in Table 3.1, not all Danish study plans and guidelines are included as sources, where no texts fulfilling the criteria for selection were found.

Data analysis

The analysis is performed at two levels: textual and discursive.

Among the several possibilities for identifying and categorising elements in the texts, I will focus on three: modality, collocation and intertextuality. Modality expresses a commitment the author makes to truth and necessity, respectively by epistemic and deontic modalities (Fairclough, 2003, p. 165). It can be evidenced – although non-exclusively – by the use of modal verbs, such as ‘may’ (epistemic) and ‘must’ (deontic). Collocation is the repeated co-occurrence of certain concepts; for example, ‘hard-working’ appears as a common binomial in political speeches. Intertextuality is the more or less explicit presence of elements of other texts and their authors’ voices. These elements can be dialogued with, assumed, rejected, and so on (Fairclough, 2003, pp. 41–42).

Beyond the mere identification of modality, collocation and intertextuality in texts, I will point out the way they are chosen to be expressed as indicators of discursive practices. In particular, I focus on legitimation strategies. Legitimation is the discursive practice of justifying what is made factual in the texts, through reference to authority, value systems or utility or conveyed through narrative (Fairclough, 2003, p. 98).

Coming back to the research questions, I define the ‘role’ of decision making as choices for the use of legitimation strategies as discursive practices. Additionally, I shall identify changes from the current to the upcoming version of the curriculum. Characterising this shift is the final stage of the analysis.

3.3.2 Selected texts

As a way of organisation, Chilean study programmes are divided into units. For each unit are described: purpose, previous knowledge, key concepts, contents, abilities, attitudes, expected learning outcomes (and their respective assessment indicators), didactical orientations, and suggested activities for each expected learning outcome. The following are the selected texts, and above them are the contexts within the study programmes where they are found. All translations from Spanish are done by me as literally as possible.

In the 3rd-middle grade current curriculum (CL315), 'decision' appears in the form of 'decision tree representations', as a follow-up for the goal of understanding the concept of conditional probability. I will not take it as part of the analysis, since it actually refers to 'probability tree representations'. Then it forms part of the general didactical orientations for the unit:

3rd-middle. Unit 4: Data and chance

Purpose

1 Experimental problems are worked with *decision* tree representations, which enable a greater understanding of content and [are] a tool for probabilistic calculations. (CL315, p. 120, emphasis added)

Didactical orientations

2 In line with this, it is fundamental that the teacher promotes the development of random thinking, i.e. that students learn to make *decisions* with evidence in situations of uncertainty. (CL315, p. 123, emphasis added)

In the last grade's current study programme (CL415) there are two units about data and chance, with no mention of 'decision' in 'data and chance 2', which includes graphic notions about binomial and normal distributions. In 'data and chance 1', 'decision' is part of the didactical orientations as in the 3rd-middle grade. Later on, there is a reference to 'decision' as a comment for teachers when engaging in activities for the learning goal to evaluate information critically:

4th-middle. Unit 3: Data and chance 1

Didactical orientations

3 In this unit, it is expected that students critically evaluate information published on the media and Internet, from the analysis, interpretation and synthesis of such information, with which they can obtain results about a population considering its size and the variable's distribution, infer conclusions from the mean, variance and standard deviation, and make *decisions* grounded in statistically significant information. (CL415, p. 86, emphasis added)

Suggested activities

> **4** Furthermore, it is important that the teacher promotes contextualised learning so students progressively develop statistical literacy, which gives them tools for making grounded *decisions*. (CL415, p. 88, emphasis added)

As for the public consultation document for the upcoming curricular framework (CL3417), decision first appears as part of the general purposes of the mathematics subject:

Mathematics: Formative purposes

> **5** In order to achieve the latter, students will work collaboratively in mathematical modelling of situations, to make grounded *decisions* in disciplinary problems, as well as in the interdisciplinary, social, environmental or economic scope. (CL3417, p. 49, emphasis added)

And then, utterances about decisions are part of mathematics learning goals in both grades:

Learning goals for 3rd-middle

> **6** It is expected that students will be capable of making *decisions* in situations of uncertainty, with information involving dispersion measures, double-entry tables and conditional probabilities. (CL3417, p. 52, emphasis added)

Learning goals for 4th-middle

> **7** It is expected that students will be capable of solving problems in contexts of uncertainty, through the application of the binomial distribution and calculation of probabilities, for *decision* making and critical analysis of statistical information. (CL3417, p. 52, emphasis added)

Both learning goals have a parallel in the previous texts. In the current study programmes (CL315 and CL415), topics such as dispersion measures, conditional probabilities and the critical analysis of statistical information are covered.

Danish study plans are published as one whole mathematics unit, and their content is organised into identity and purpose; subject goals and content; organisation; and evaluation. Guideline documents have the same structure and refer to the study plans in a more detailed and comprehensive manner. All texts were translated to English with Google translator and revised by a Danish colleague.

Making decisions is first mentioned as part of the general purpose of mathematics in high school in a study plan (DKB17) and guideline document (DKAB18), in regard to what students are enabled to do:

Purpose

> **8** The work on mathematical material leads the student to gain knowledge and skills within mathematics and enables the individual to understand, analyse, evaluate and make *decisions* in social, professional and study contexts. (DKB17, p. 1, emphasis added)

Purpose

9 In addition to gaining knowledge and skills within mathematics, the individual student is enabled to understand, evaluate and make *decisions* in everyday, professional and study contexts. (DKAB18, p. 2, emphasis added)

In a more specific reference to statistics, the relevance of data-driven decisions in policy is drawn in two guideline documents (DKBC18, DKABC18):

Subject goals and content – Core material and minimum requirements – Statistics

10 With the development of computers, the Internet and means of communication, amounts of data are now stored in an order of magnitude that is growing explosively, and statistics based on data is increasingly being used in political *decisions*. (DKBC18, p. 8; DKABC18, p. 8; emphasis added)

The one utterance about choice is contained in the Danish statistics experts' document, in the context of explaining the importance of learning statistics.

Why is it important to learn statistics?

11 This applies, for example, to the Monty Hall problem (Ellis, 1995), where a participant on a TV show has to *choose* between three doors behind one of which there is a car. (DKSG17, p. 2, emphasis added)

The above 11 sentences are the units of analysis for the following textual analysis and identification of discursive practices. Narrowing down to isolated sentences is an evident limitation of the proposed analytical strategy, and I address it by two means. First, sources and headlines (in italic font) are considered in the analysis as being part of the texts, as an explicit acknowledgement of the context they are written in. Second, textual and discursive elements of the analysis take other parts of the source documents when perceived necessary.

3.3.3 Textual analysis

I shall first identify and describe elements of the textual analysis found in the excerpts to provide input to the following discussion. In other words, in this section, I extract what is in the texts, before arguing their meaning. I focus on modality, collocation and intertextuality.

Modality

Starting with the Chilean texts, modal forms expressed as 'it is expected that students' (3, 6 and 7) can be identified as epistemic modalities, i.e. as expressions of probability and truth, in this case, the expected and not certain to happen. It can be argued that, given the official character of curricula, these are actual expressions of the necessary, falling into the category of deontic modalities.

Explicit deontic modalities are evidenced as 'it is fundamental that' (2) and 'it is important that' (4), and they express the necessity for particular attention to be paid by the teachers in enabling students' decision-making skills.

Danish extracts express modalities more implicitly. Except for the one modal form ‘has to choose’ (11), which refers to the rules of a game show, deontic modalities are exposed in the sections they are written. The mention of the Monty Hall problem exemplifies, in part, *why it is important to learn statistics* (11). Likewise, texts (8) and (9) are part of the subject’s *purpose*, so it is desired for the student to ‘gain knowledge’ and ‘be enabled’ (8, 9).

Collocation

References to decisions do not appear alone. A habitual co-occurrence of the substantive ‘decision(s)’ in Chilean texts comes with ‘grounded’ as a company, both as an adjective as in ‘grounded decision making’ (4) and ‘grounded decisions’ (5), and as an adverbial form, as in ‘decisions grounded in…’ (3). A similar adverbial accompanying form is ‘decisions with evidence’ (2) and ‘decisions with information’ (6). This collocation suggests a reference to a particular type of decision or decision-making processes, based on quantitative arguments, distinguishable from a mere act of making a choice.

A different collocation appears in Danish texts. Decisions are accompanied by their contexts, whether they are ‘in social, professional and study’ (8), ‘everyday, professional and study’ (9), or ‘political’ (10). It is also the case for the verb ‘to choose’ written together with its particular context of a TV show (11).

Intertextuality

Intertextuality is found within selected texts. In the Chilean case, texts 2 and 4 can be read in parallel as having the same structure:

In line with this,	Furthermore,
it is fundamental that the teacher promotes	it is important that the teacher promotes
	contextualised learning so students
the development of random thinking	progressively develop statistical literacy
i.e. that students learn to make decisions with	which gives them tools for making
evidence in situations of uncertainty. (2)	grounded decisions. (4)

These texts make references to two different concepts, namely ‘random thinking’ (2) and ‘statistical literacy’ (4) in a similar way: as notions to be promoted by the teacher, driving students to make grounded decisions.

From the Danish sources, texts 8 and 9 also share structure and content. As much as guideline documents refer to the corresponding study plans, text 9 is not an actual quote from text 8:

The work on mathematical material leads	In addition
the student to gain knowledge and skills	to gain knowledge and skills
within mathematics,	within mathematics,
and enables the individual to understand,	the individual student is enabled to understand,
analyse, evaluate and make decisions in	evaluate, and make decisions in
social, professional and study contexts. (8)	everyday, professional and study context. (9)

It is no surprise that a text from the guidelines refers to the corresponding study plan, but more interesting are the subtle differences. From the B mathematics programme (8) to the guidelines for A/B mathematics (9), the ability to 'analyse' decisions is cut out. Moreover, decisions in 'social' (8) contexts are exchanged for 'everyday' (9) context.

There are also references to texts other than those selected. 'Random thinking' (2) and 'statistical literacy' (4) are not defined in the documents, but they are traces of another text published by the Chilean Statistics Society as experts' curricular recommendations (Araneda, del Pino, Estrella, Icaza, & San Martín, 2011). To the best of my knowledge, this is the one Chilean document containing both concepts together as key necessary goals of learning statistics in high school. The term 'random thinking' is not found in the literature. Rather I presume the intention is to refer to 'statistical' (Garfield & Ben-Zvi, 2007), or 'probabilistic thinking' (Chernoff & Sriraman, 2014b; Borovcnik, 2017). Both 'statistical thinking' and 'statistical literacy' are described as core answers to section *Why teach statistics?* (Araneda et al., 2011, pp. 1—17), referencing a paper published in the International Statistical Review (Garfield & Ben-Zvi, 2007). Overall, texts 2 and 4 are referring to authority, introducing notions without further explanation, which are developed within national and international statisticians' associations. In Danish texts, the one and explicit reference to an external text is a mention of 'the Monty Hall problem' (11) by citing a website where it is explained. It is not an academic or official reference, and the website does not provide information on the author's credentials. The relevance of this remark lies in the problem itself being a classic example of how intuition and a probability-based rationality lead to opposite courses of action. This statement is the core of the research agenda of behavioural economics and rational choice (Kahneman, 2011), and it has been proposed in statistics education research as a resource to address probabilistic misconceptions (Batanero et al., 2009; Elicer & Carrasco, 2017).

3.3.4 Discussion: Discursive practices

Within Fairclough's (1992) three-dimensional model, texts are part of social events, and they are signs of discursive practices. Discourse mediates the relationships between texts and social structures. In the following, I point out and problematise representations of social actors and legitimation strategies evidenced by the preceding elements of the textual analysis.

Representation of social actors

As a whole, and given the genre, texts include three generic and classified social actors: students; teachers; and politicians. The first two are, of course, the case of upper secondary school students and teachers. In different texts referring to decisions, these actors are also excluded via suppression and backgrounding (Fairclough, 2003, p. 145).

Teachers are suppressed from the current (CL315, CL415) to the future Chilean curriculum (4M17). In 2 and 4, teachers' actions are the ones triggering students' decision processes in a certain way, whilst in 5, this is engaged through students' collaborative work. In hand with this suppression of teachers, students' presence changes from passive (as affected by teachers' actions) to active. Moreover, a key shift in the formulation is evidenced by the fact that students' (or future citizens') decisions are present not only in declarations of purpose (1, 5), didactical orientations (2) and suggested activities (4), but as part of the expected learning outcomes. In particular, students are expected to be capable of 'making decisions' (6) and 'solving problems (. . .) for decision making'. The shift in Chilean texts is coherent with the Danish case. Students' ability to make decisions (8, 9) is a result of the 'work on mathematical material' (8), leaving teachers' participation implicit.

The collocation of decisions to their contexts in Danish texts indicates a scope that Chilean texts do not hold, making students able to decide in 'social' (8), 'everyday' (9), 'professional and study' (8, 9) contexts. It must be noted that these texts come from the generic mathematics programme, as opposed to specific 'data and chance' educational units. In fact, Chilean texts 6 and 7 reflect an alteration that is coherent with another broader reference to 'decision' in the diagnosis chapter of the same document: "They demand protagonism in decision making and aspire to contribute to solving problems in the world they live in, such as poverty eradication, climate change and sustainable development" (MINEDUC, 2017, p. 13).

The wish for future citizens' participation in decisions under uncertainty goes in hand with Ulrich Beck's observation about the role of experts in the risk society. Experts point out risks, and the people decide to take action in a complex political process. However, "how one acts in this situation is no longer something that can be decided by experts" (Beck, 2000, p. 217). Both countries push for an increasingly active role of students in deliberation as part of the purposes of school mathematics. However, it can be argued that the specificity of probability and statistics has not yet been addressed.

Legitimation

Making decisions appears as a way of justifying the teaching and learning of statistics in the curricula. The texts fulfil the purpose of not only saying what and how to teach and learn but also why and for what. Fairclough (2003, p. 219) claims that much of the legitimation of a social order – such as the inclusion of particular knowledge in the official curriculum – is textual. In particular, the texts show three of the legitimation strategies identified by van Leeuwen (2007), namely through mythopoesis, rationalisation, and authorisation.

Mythopoesis or legitimation conveyed through a narrative is evidenced by the collocation of 'decision' with 'grounded' (3, 4 and 5), and similar accompanying adverbial forms such as 'with evidence' (2) and 'with information' (6) in Chilean texts. This strategy is stable through the texts and makes a case for justifying probability and statistics since not just any kind of decision is included, but only those rooted in data and mathematical rationality. It allows a steady association between such school subjects and rational choice. That is not the case for Danish texts, where decisions are instead collocated with a variety of contexts, 'political decisions' being the only ones said to be made using 'statistics based on data' (10).

Merely associating probability and statistics to rational decisions is not enough. It is still necessary to justify their teaching. Aforementioned deontic modalities are evidence of legitimation through rationalisation, which refers to utility. Within the current Chilean curriculum (CL315, CL415), teachers' promotions are 'fundamental' (2) and 'important' (4) for students to make grounded decisions. The upcoming proposal (CL3417) provides a similar rationalisation in the form of 'in order to achieve the latter' and 'to make grounded decisions' (5), but this time it is mediated by the students' actions, namely their collaborative work in mathematical modelling situations (5). In a way, this change represents another shift in the agency from teachers to students in the teaching-learning process. In a similar but implicit manner, modalities in Danish texts reflect that part of the purpose of the ‘work on mathematics’ (8, 9) is to lead the student to make decisions, and that ‘it is important to learn statistics’ (11) to address choices obscured by intuition.

As already suggested by intertextual elements, traces of an authorisation strategy are found in texts from the current Chilean curriculum (CL315, CL415), through the inclusion, without further definition, of ‘random thinking’ and ‘statistical literacy’, and the Monty Hall problem in the Danish case. I highlight the fact that the authorisation strategy is not found in the Chilean upcoming curricular framework (CL3417). This shift resonates with the bottom-up nature of the latter document, where civil society plays a more significant role in the justification of the curriculum. A difference worth noting is that, while Chilean curricula do not make explicit reference to statisticians’ recommendations (Araneda et al., 2011), the Danish Ministry of Education actually endorses and publishes experts’ impressions, and so they are taken as a source in this article (DKSG18). Moreover, the collocation of ‘political decisions’ (10) as a case for the relevance of learning statistics defines a clear call for authority; if ‘statistics based on data is increasingly being used’ (10) by political authorities, students and future citizens ought to participate in statistical deliberation.

3.3.5 Summing up

My overall intention here is to investigate the discourses of decision making within official curricular documents regarding probability and statistics as school mathematics content. The goal is to exemplify possible links between mathematics education and deliberative democracy, a concern shared by the Nordic and Latin American communities.

Critical discourse analysis allows me to see texts as part of social structures mediated by discursive practices, which give meaning to what is in the texts, operationalising the search for meaning pragmatically and systematically, in order to address two research questions.

The first research question intends to reveal a diversity of discourses attached to notions of decision making as expressed in Chilean and Danish upper secondary school curriculum. It is possible to evidence strategies for positioning social actors and legitimation. Decisions position actors depending on who is making them. Students are encouraged to participate in such processes either by teachers’ actions or by their work in school mathematics. Present legitimation strategies show that official study programmes not only describe what is to be taught but also for what purposes. Decision making appears as a central element of such legitimation, by establishing a narrative of probability and statistics linked to grounded, evidence-based and rational choice, and the teaching of those subjects as necessary and relevant. Legitimation appeals to political authorities, professional statisticians and statistics education researchers’ voices.

The second research question refers to contrast, and though curricular texts selected from both countries reveal similar discursive practices, they perform them differently. The analysis illustrates signs of change towards a justification that relies less on professional statisticians and educational researchers as authorities in Chile, and more on students’ learning activities as decision-makers. This shift is part of the wish for a more significant role in decision processes that go beyond the disciplinary scope. In the Danish case, students’ active role is already stressed in the curriculum.

Chilean curricular excerpts reveal a persistent legitimation of the learning of probability and statistics conveyed through narratives of rational decision making. Instead, Danish texts tend to attach decisions to a diversity of contexts, from the general political and social, to explicit mentions of games and school subject matter. There is a common appeal to statisticians and statistics education researchers within the psychological and economic theories of rational choice, the Danish case being more explicit than the Chilean. Despite the distinctions, a key finding in this respect is that both countries exemplify the purpose of mathematics education to enable students to participate in societal and political decision-making processes, but fail to specify the relevance of probability and statistics in such contexts, relegating their scope to gambling games and other individual rational choices.

3.4 Narrowing down by broadening up the scope of decisions

Let us come back to the justification problem by outlining relevant reasons and aims for the teaching of probability and statistics in upper secondary school, through the notion of decision making. The thesis's general *problématique* concerns the connection to teaching practice, and therefore I focus on how these justifications are depicted in curricular documents. On that same line, given that the empirical basis of the thesis takes place in Chile, those texts will be the centre of attention.

Niss (1996, p. 13) suggests that reasons for mathematics education can fit into three overarching categories. The first one is for "contributing to the technological and socio-economic development of society at large". The second is for "contributing to society's political, ideological and cultural maintenance and development". These reasons may refer either to society at large or as a competition with other societies or countries. The third reason is to provide "individuals with prerequisites which may help them to cope with life in the various spheres in which they live; education or occupation; private life; social life; life as a citizen".

Reasons within the first type are not quite elicited in the texts analysed. That is not to say that they do not exist at all. For instance, Araneda et al., among their answers to *why to teach statistics* in Chile, they state that:

> A statistically competent workforce will allow the country to compete more effectively in the world market economy and to improve its position in the international economic market. For example, statistics provides tools for achieving improvements in quality and productivity that such competitive market demands. An investment in statistical literacy is thus an investment in the economic future of our nation."
>
> (Araneda et al., 2011, p. 13)

From that perspective, the inclusion of data and chance in the Chilean mathematics curriculum becomes relevant for the economic sake of the nation. For certain, this is not exclusive to statistics, let alone Chile. This national economic perspective is a clear example of what Valero (2017) problematises regarding justifications for *mathematics for all*. As a theme, it has become a self-evident truth in a bit more than a century. This justification is grounded in quantitative studies correlating mathematics achievement in standardised tests with countries' economic growth, and competitiveness (Popkewitz, Diaz, & Kirchgasler, 2017). Valero argues that "all these threads of discourse operate under a similar logic of human capital management through measurement of educational achievement" (2017, p. 21). The problematisation is not whether school mathematics achievement and economic progress are linked, but the fact that reducing citizens to human capital can be threatening to democratic values (Brown, 2015).

The historical development of probability and statistics as a mathematisation of decisions under risk and uncertainty can be illustrative of the second type of reasons for their inclusion in the curriculum. As seen in Section 3.1, making, rationalising and mathematising decisions have been a driving force for the development for probability and statistics, and a reason for their inclusion in the curriculum may well be the maintenance and development for the maintenance and development of such practices. The analysis in Section 3.3 shows that the legitimation of probability is conveyed through narratives of rationality, appealing to psychological and economic theories of rational choice, even making the explicit case of chance games. Appealing to the historical development of probability embedded in social practices of decision making under uncertainty is an ongoing research agenda, from a socioepistemological perspective. Vergara Gómez and Espinoza Ramírez (2018) claim that this may allow redirecting and re-signifying school probability. As for statistics, curricular texts legitimise its inclusion as being a commonly used tool for political decisions, and hence being necessary for the continuation of the practice of social statistics (Donnelly, 1998).

At large, the texts analysed refer to reasons in the third category, by arguing for what the individual students are enabled to be and do, through probability and statistics. The genre of the sources can explain this. Study programmes are directed to teachers to work with their group of students. Decisions in a variety of contexts are invoked in curricular texts, both in Chile and Denmark (Section 3.3). Individual economic-driven decision scenarios have been studied in the heuristics and biases programme related to probabilistic thinking (Batanero & Borovcnik, 2016). As seen in Section 3.2, heuristics and consequential errors driving us away from this rationality have informed probability and statistics education research on the possible pitch falls, difficulties and the complexity of contextualised decisions. However, these are persistent and formal education and training have not proven yet to make students or event experts less prone to systematic errors. Decisions in private life with the use of statistics are present in curricular texts as well, claiming to provide students with tools for grounded and informed decisions. In that regard, statistical literacy and the use of statistically significant information are brought up to serve that purpose in Chilean texts. This is the perspective Wild et al. (2018) take when they pose the question "why learning statistics is more important than ever?" For them, statistics education scholars have the responsibility of "figuring out how best to convey these ideas in ways that are useful, and that will allow students to make better decisions using data" (2018, p. 27). Here, both probability and statistics in school are justified to be part of the school curriculum for they provide students with tools to cope with situations of uncertainty and data-richness, in order to make rational decisions for their best sake.

3.4.1 Where is the critical part?

Let us zoom out and consider probability and statistics as part of mathematics, at least as school subjects. In particular, given the anchoring to decision making analysed in the texts, let us consider them as applied mathematical subjects. A critical justification for the inclusion of modelling and applications of mathematics in school can be defined as follows:

> The *critical competence* argument focuses on preparing students to live and act with integrity as private and social citizens, possessing a critical competence in a society the shape and functioning of which are being increasingly influenced by the utilisation of mathematics through applications and modelling. The aim of such a critical competence is to enable students to *see and judge* independently, to recognise, understand, analyse and assess representative examples of actual uses of mathematics, including (suggested) solutions to socially significant problems.
>
> (Blum & Niss, 1991, p. 43)

The premise of the justification is that the world is shaped and influenced by applications of mathematics, i.e. it presumes that mathematics has a formatting power in society (Skovsmose, 1994; Niss, 1994). Provided that assumption, the aim is to enable students to recognise, understand, analyse, assess and even suggest solutions to societal issues. More recently, instead of *justifications*, Kaiser, Blomhøj, and Sriraman (2006) refer to different *perspectives* of modelling, being the one above called the socio-critical perspective. From this point of view, "critiquing modelling is part of the learning that takes place in the process of doing modelling, and one of the aims is to produce critical, politically engaged citizens" (Barbosa, 2006, p. 296).

Do curricular texts resemble this justification or perspective for probability and statistics? At least concerning the premise, curricular texts do go beyond the personal scope. Socially relevant political decisions are invoked as well. However, if social and political decisions are among the justifications of probabilistic and statistical ideas in the curriculum, evidence of such empowerment is needed to make that connection more transparent. As opposed to the notion of *powerful mathematical ideas* (Skovsmose & Valero, 2002), probability and statistics are perceived as relevant for being *mathematical ideas of the powerful*. It is said that political authorities and certain professionals make use of this knowledge to make socially relevant decisions, but it is not clear if and how it empowers them. Moreover, given that most students will certainly not be in those positions of power, the empowerment of all citizens through probability and statistics is yet to be thoroughly discussed. Overall, texts call for students to make rational decisions in private life and the mention of broader decisions in MINEDUC (2017) is not specific to mathematics, let alone probability or statistics. By addressing **RQ1** and **RQ2**, I can establish that the curriculum is part of the justification problem within the *problématique* of critical mathematics education, because, at most, only accepts the premise of the critical competence argument. This omission is present in the research literature as well. In his commentary on the psychological perspective of probabilistic thinking, Greer (2014) highlights the relevance of probability and statistics concerning social and political issues, and urges to go beyond providing citizens with the probabilistic and statistical tools for making their own decisions in an uncertain world for their sake, as implied by Meder and Gigerenzer (2014); but to address social and political implications of decisions onto that world. For example, for discussing the diverse issues of a probability literacy for all citizens, Gal (2005) uses a fictional conversation between a man and his mother to decide whether he should get health insurance. The discussion is rich in the language of data, proportions, chances and probabilities, but at no point, they question the social and political construction behind and implications thereof a privatised health insurance system. With Greer, my normative take is that for socially and politically relevant decisions to make sense as being part of the justification, exploring how probability and statistics shape the world ought to be explored in upper secondary school mathematics, the educational instance to which all future citizens are entitled. Making that connection is the ambition of what follows in this thesis.

I have provided a broad landscape of justifications for the teaching and learning of probability and statistics in upper secondary school, grounded in the historical development of the subject matter, perspectives in research literature agendas, and illustrated in curricular documents. I have sorted out that a critical justification is not strongly stated in the Chilean curriculum. The following part of the thesis is to see whether it is still possible to design and implement a teaching practice compatible with that justification, through four classroom experiences. Though each learning environment is distinct in the mathematical content, grade and contextualisation, there are some general guidelines for approaching them, which I detail on Chapter 4. As a point of departure, in order to walk along the line of the critical justification, among the many situations where decisions take place, I will appeal to socially relevant implications of decisions. That is, by broadening up the scope of decisions from personal to social, I let myself narrow down the scope of the thesis.

Notes

[1]Not to be confused with middle school, as in between primary and secondary school. In Chile, *enseñanza media* (middle education) is the bridge between *enseñanza básica* (basic education) and *educación superior* (higher education).

[2]The source CL3417 was used for the analysis in its version up to May 2019. However, by the publication date of Elicer (2019) and Elicer (2020) and let alone the writing of this thesis, it has been updated by MINEDUC, from a document for public consultation to an official curricular framework (MINEDUC, 2019). The formulation of text 7 has been altered to:

> It is expected that students will be capable of grounding decisions in situations of uncertainty, from the critical analysis of statistical data and based on binomial and normal models. (MINEDUC, 2019, p. 111)

Here, students are not expected to 'make' decisions anymore, but to 'ground' or 'justify' them. This shift is of utmost relevance for future analysis.

Chapter 4

Methodological generalities

In Chapter 3 I have dealt with the role of decision making in the justification of probability and statistics in mathematics education, embedded in upper secondary school mathematics curricula. The analytical strategy is unique to this part of the research project, where the theoretical framework, collection and analysis of data are inspired in Fairclough's Critical Discourse Analysis (e.g. 1992, 2003). Other methodological approaches are needed to move along the thread of coherence of the critical justification within teaching practices. These do not encompass the analysis of curricular texts, but the design, implementation and analysis of classroom activity.

The critical justification, intended to be experienced, has a broad reach. It aims for the inclusion of notions of probability and statistics, devoted to 11th and 12th-grade students for the Chilean case, and invoking decisions and applications to socially relevant issues. For this reason, I have decided to explore possibility and implementation problem fields of the *problématique* in four different tasks, each of which comprises distinctive aspects of the subject matter, contextualisation of the activities and analytical perspectives. Nevertheless, some aspects for the design, collection and analysis of data apply to all, which therefore deserve a prior clarification.

4.1 Critical motives, interpretivist paradigm

As discussed in Chapter 2, CME can be portrayed as a set of concerns of issues of power in mathematics and mathematics education. In this thesis, I focus mainly on the formatting power of probability and statistics. However, the research project is not embedded in a critical research paradigm. According to Guba and Lincoln (1994, p. 113), critical research can be characterised by its aim:

> The aim of inquiry is the critique and the transformation of the social, political, cultural, economic, ethnic and gender structures that constrain and exploit humankind, by engagement in confrontation, even conflict. (...) The inquirer is cast in the role of instigator and facilitator, implying that the inquirer understands *a priori* what transformations are needed.

Skovsmose and Borba (2004) argue that such a definition implies that the researcher understands the problems and transformations needed and plays the role of leader, which I do not pretend to. The focus in my research project is whether the participants, teachers and primarily students, can engage in critique through a teaching practice of probability and statistics. Therefore, critical inquiry provides content to my research questions, but not their formulation. In that sense, the methodological approach needs not to *be*, but to *resonate* with the concerns of critical mathematics education in the classroom (Skovsmose & Borba, 2004).

Instead, the overall research paradigm of this thesis is interpretivist. Interpretivism arises in the social sciences as an alternative to scientific positivism and "is founded upon the view that a strategy is required that respects the differences between people and the objects of the natural sciences and therefore requires the social scientist to grasp the subjective meaning of social action" (Bryman, 2015, p. 26). My research questions seek to describe and understand students' engagement in the classroom activities and speculate possibilities. The exercise of explaining how students' handle activities and what critical issues are there illustrated cannot be objectified. A claim of the kind "a group of students is reflecting" is open to interpretation by the researcher. It is also subjected to the social, educational, and experimental context that will be described in the methods and analysis.

On the same line of reasoning, I have chosen a qualitative research strategy, in the sense of looking for meaning in students' work and discussions and let that meaning inform the general *problématique*. My intention is to address possibilities in educational practice, researching not only what *is* the case but what *could be* the case (Skovsmose & Borba, 2004; Blomhøj, 2006). Therefore, quantitative strategies are not relevant to my research questions. Whether one or a majority of students engage in critical reflections during the classroom experiences is not as relevant as the content within. However, an evident challenge of external validity arises, as it is the case for many qualitative strategies, whereby

> rather than generalising from a random sample to a population (statistical generalisation), many (mainly qualitative) research approaches aim for generalisation to a theory, model or concept (theoretical or analytic generalisation) by presenting findings as particular cases of a more general model or concept (Yin, 2009).
>
> (Bakker & van Eerde, 2015, p. 434)

Bakker and colleagues speak from the standpoint of design-based research (see, e.g P. Cobb et al., 2003; Design-Based Research Collective, 2003), where the cyclic nature of the strategy aids its aim for generalising a design as case study.

As a way to tackle the issue of external validity, I examine the general *problématique* through four classroom experiences. Hence, the methodological design of this thesis can be characterised as a *multiple-case study*. One evident advantage to this design is that by comparing cases, one is in a better position to discuss the circumstances where a theory holds (Yin, 2009). More importantly, for this thesis, Bryman (2015, p. 67) argues that " the comparison may itself suggest concepts that are relevant to an emerging theory". The critical lens in probability and statistics education is a young and ongoing research programme and calls for different, and less direct approaches (Petocz, Reid, & Gal, 2018).

Each of the four classroom experiences defines one case, that can be classified as being *representative, typical* or *exemplifying*, as Bryman (2015) calls it. When choosing exemplary cases, "the objective is to capture the circumstances and conditions of an everyday or commonplace situation" (Yin, 2009, p. 48). Accordingly, as the general *problématique* of the thesis is concerned about a teaching practice under genuine circumstances, the effort is put into researching regular Chilean upper secondary school classrooms. Consequently, in the quest for teachers and upper secondary schools to provide the cases, I decided to decline two offers. One of them was from a private school of a high-class neighbourhood of the capital. The other case involved a teacher who was, in fact, a business school graduate, with no previous teaching experience and no intention to follow after two years. Instead, two of the schools involved in this thesis are either public or subsidised, and all three teachers are professional mathematics teachers. Though Bernardo's school is private, its location is meaningful for the environmental struggle illustrated in the task. Their circumstances are representative of the Chilean school and social landscape.

A summary of the schools involved in the thesis is given in Table 4.1. The source of all information is the Chilean Agency for Quality in Education, collected at the time students were in 10th grade, i.e. 2016 for 11th-grade classes and 2017 for 12th-grade classes (Agencia de Calidad de la Educación, 2016, 2017, 2019).

Ch.	Grade	Teacher	Type	Socioeconomic level[3]	Description[4]
5 6	11th 12th	Alejandra	Subsidised	Low	Most parents or tutors declare up to 8 years of education for the father and 9 for the mother. Between 94.01% and 100% of students are found in condition of social vulnerability.
7	11th	Bernardo	Private	High	Most parents or tutors declare 16 or more years of education. 6% of students or less are found in condition of social vulnerability.
8	12th	Carla	Public	Middle	Most parents or tutors declare 11-12 years of education for the father and 12-13 for the mother. Between 85.01% and 94% of students are found in condition of social vulnerability.

Table 4.1: Schools involved in the PhD project

As for the specific research methods, a thorough description follows. In Section 4.2 I delineate three design principles for the tasks, namely *exemplarity*, *inquiry approach* and *pragmatism*. I briefly describe the four tasks implemented and how they are preliminarily coherent with the design principles. Other than supporting the choice for cases and design of the tasks, these principles, their interpretations and tensions in-between will be in constant revision throughout the implementation, in order to address **RQ3**. In Section 4.3 I describe the method for producing data, providing details on how the interventions came to be, the role of different participants, and the collection of classroom data. Joining the notion of teachers as research subjects more than objects in critical mathematics education research (Skovsmose & Borba, 2004), in Section 4.4 I take a further step describing semi-structured interviews with two of the three teachers involved in the project. These interviews have two purposes; to enrich the analysis of classroom data, thus digging into what *did* happen; and to open up perspectives for further teaching practice in light of the *problématique* of coherence, thus imagining what *could* happen, anchored in the classroom experience. In Section 4.5, I describe how the analysis is carried out throughout the thesis. Addressing research questions **RQ4** and **RQ5** implies, firstly, to observe and make sense of critical reflections on behalf of the students. Secondly, adherence to the critical justification implies identifying aspects connected to critical reflections which are specific to probability and statistics.

4.2 Design of learning environments

In the realm of probability education, there is a vast body of research in qualities of teaching probability (Sharma, 2016). A clear trend can be observed as a distancing from so-called "traditional" approaches, i.e. the vertical transmission of knowledge from teacher to students in a mono-logical form (Alrø & Skovsmose, 2002). For instance, Vahey, Enyedy, and Gifford (2000) developed an inquiry-based simulation environment with middle school students, reporting that the participating group outperformed the control group following a traditional curriculum in understanding probability concepts. A similar result is reported by the cross-age comparative study by Gürbüz, Erdem, and Fırat (2014), showing how an activity-based teaching experience, with focus on an argumentation-safe environment, played a significant role in students' learning of some probability concepts. Prediger and Rolka (2009) highlight the role of collaborative communication in the construction of agreed strategies when dealing with betting board games amongst 12-year-olds. More recently, by reviewing diverse views on effective probability teaching practices, Sharma (2016) develops a learning environment to explore the fairness of a dice rolling game, where "the sequence starts with a problem, rather than with mathematical theory or procedure, and the progression is from the empirical to the theoretical"(Sharma, 2016, p. 134).

The studies mentioned above are focused mostly on the purpose of developing probabilistic concepts and thinking. The chapter by Greer and Mukhopadhyay (2005) can be considered a seminal work on the teaching and learning with historical, cultural social and political contexts in mind. They propose three directions to move forward, which are undoubtedly relevant for this thesis: "deepening of historical, epistemological, and cultural analysis, strengthening the links between probability as taught in school and the lives of students, and education for probabilistic modelling as a tool for critical analysis of social and political issues" (Greer & Mukhopadhyay, 2005, p. 314).

As for statistics education research, the development of general principles for the design of statistics learning environments is summarised by P. Cobb and McClain (2004) and revisited by Ben-Zvi, Gravemeijer, and Ainley (2018). Their work proposes seven dimensions to be aware of: the focus on developing central statistical ideas rather than on tools and procedures, the use of well-designed tasks to support the development of statistical reasoning, the use of real or realistic and motivating data sets, the use of technological tools that allow analysing data, the establishment of a classroom culture that fosters statistical arguments, and the use of assessment. These principles and their interconnections are a useful broad view on task design, though they are focused essentially on the purpose of developing statistical concepts and thinking. However, "researching the inner change process that learners' undergo as they develop the 'critical lens' that is part of statistical literacy requires different and less direct approaches" (Petocz et al., 2018, p. 81), such as open-ended investigations. Recent studies echo this idea. For example, by asking openly to interpret a chart relating school recess time with race, (Brantlinger, 2014) students engage spontaneously in discussions about systemic racism, on the grounds of graphical representations of statistical information. Kuntze, Aizikovitsh-Udi, and Clarke (2017) show it is possible to design hybrid tasks that provoke both statistical or critical thinking, through a "thinking-aloud" task consisting of the evaluation of a claim based on a diagram of births and deaths in Germany since 1945. These studies make use of real data from sensible contexts with problematic representations and a variety of possible interpretations. As a critical improvement to the tasks, the authors propose a reflection-oriented framing, i.e. to have the learners ask themselves critical questions after solving the task.

In order to address **RQ3**, I combine the ideas of task design into three overarching principles: *exemplarity*, *inquiry approach* and *pragmatism*, which I describe in the following. These principles not only facilitate the choices for contexts and practices to apply in the classroom interventions, but they define the general dimensions to focus the ex-post analysis in Chapter 9.

4.2.1 Exemplarity

As a pedagogical concept, the exemplary principle can be traced back to science educator Martin Wagenschein. He introduced the concept to capture the idea that it is possible and necessary, in institutionalised education, to teach and for the students to learn about general phenomena through in-depth studies of a few well-chosen cases. Although some researchers present Wagenschein's position mainly as a "method of reducing curricula" (Andersen & Kjeldsen, 2015, p. 25) in such a way that "the partial serves as a mirror of the whole" (Wagenschein, 1956, pp. 133-134), his call for courage to allow holes in the curriculum must be justified by arguing how the general reveals itself through the specific. The exemplary principle, then, does not refer to the mere act of case-teaching, but rather to the quality of the case. The question is not so much about the specificity or simpleness of the activity. It is about what it is exemplary for. Three general perspectives on exemplarity in education are relevant to this thesis, namely subjective, instrumental and critical.

The first tradition of exemplarity can be connected to experiential education in general pedagogy as proposed by John Dewey, i.e. as a source of motivation for learning. The orientation is towards the learner's **subjectivity**. Illeris (2002) draws on Dewey to explain how an exemplary problem has its roots in connecting the subject matter to the students' experienced life, because

> ... an experience arouses curiosity, strengthens initiative, and sets up desires and purposes that are sufficiently intense to carry a person over dead places in the future ... [and not] operate so as to leave a person arrested on a low plane of development, in a way which limits capacity for growth.
>
> (Dewey, 1938/2015 in Illeris, 2002, p. 148)

From this perspective, the personal relevance to a student defines an example's quality, since it can sustain her energy and motivation to go through a learning process that can often be challenging.

Within the same didactical perspective, exemplarity is **instrumental** to disciplinary concepts and methods. A case is exemplary of the language and ways of the discipline; it ought to cast light over the whole field. Wagenschein himself speaks from the viewpoint of the natural sciences, arguing for exemplary problems to be illustrative of the fundamental laws of physics. He claims that all disciplines have elementary problems and questions and that students can get insight into the whole field by working on one of these elementary problems. This perspective is close to what, in the context of teaching mathematical models and applications, Niss and Blum label as "modelling for the sake of mathematics (learning and appreciation)" (2020, p. 28), though in teaching *pure* mathematics, exemplary problems need not be contextualised applications, such as in number theory, algebra, geometry, and even (mathematical) probability theory, to some extent.

However, this is not the case for statistics education, as it is acknowledged in the GAISE report (Franklin & Garfield, 2006; Franklin et al., 2007), endorsed by the American Statistical Association:

> Many mathematics problems arise from applied contexts, but the context is removed to reveal mathematical patterns. Statisticians, like mathematicians, look for patterns, but the meaning of the patterns depends on the context.
>
> (Franklin et al., 2007, p. 7)

Therefore, the statistical disciplinary realm cannot be detached to context as mathematics can, for "statistics requires a different kind of thinking, because *data are not just numbers, they are numbers with a context*" (G. W. Cobb & Moore, 1997, p. 801, emphasis in original).

From an instrumental perspective, contextualised examples should be exemplary of statistical disciplinary concepts. This attachment defines one of the six design dimensions for statistics learning environments (P. Cobb & McClain, 2004; Ben-Zvi, Gravemeijer, & Ainley, 2018), where the focus ought to be put on "developing central statistical ideas rather than on tools and procedures" (Ben-Zvi, Gravemeijer, & Ainley, 2018, p. 488). These central ideas include data, distribution, centre, variability, comparing groups, sampling, modelling, inference and co-variation (Garfield & Ben-Zvi, 2008). As for big ideas about probability, beyond its elemental connections to statistical ones (e.g. probability distributions and fundamental theorems that enable inference), Borovcnik claims that "little has been published on an explicit description of probabilistic thinking and the mathematical competencies that are related to it" (2017, p. 1496). That is not to say that insufficient work has been done in describing elements of probabilistic thinking (e.g. Chernoff & Sriraman, 2014b), but an established categorisation of big probabilistic ideas is an ongoing task. In the context of risk, Borovcnik (2017, sec. 2.2) synthesises ideas of Batanero and Borovcnik (2016, p. 20) into three: the theoretical character of probability and independence, conditional probability, and probabilistic evidence.

The other major perspective, namely the **critical**, is oriented towards action and empowerment. It can be associated with German philosopher and sociologist Oskar Negt. He reformulated the exemplary principle in the context of workers education as a pedagogical approach that can provide the working class with "possibilities for self-liberation" (Negt, 1975, p. 42). For him, these can only be enabled by an understanding of the power structures of society, and of how these structures create subjective and objective conditions for repression and exploitation of workers. Negt's notion of exemplarity draws from C. Wright Mills' notion of *sociological imagination* (Mills, 1959/2000), described as "the ability to switch 'from one point of view to another (...) from the political to the psychological, from the examination of the individual family to the evaluation of the state budget' and to realise structural connections between individual life stories, immediate interests, wishes, hopes and historical events" (Negt, 1975, p. 45). Thereby, the quality of an example is its ability to connect and illustrate broader political context and history. In this sense, a mathematics educational practice can be exemplary of the formatting power of mathematics (Skovsmose, 1994, Chapter 3), going beyond mathematical and technological knowing, but to reflective knowing (Skovsmose, 1994, Chapter 6). Mathematical knowing refers to internal consistencies in calculations, and technological knowing to the appropriateness of the mathematical approach put into context. Enabling reflective knowing, however, implies prompting to ask whether mathematics is needed at all, what are the broader consequences of the approach taken, and about our own and alternative ways we are reflecting. A parallel can be made to what Weiland refers to as "making sense of language and statistical symbols systems and critiquing statistical information and data-based arguments encountered in diverse contexts to gain an awareness of the systemic structures at play in society" (2017, p. 41). As for probability, study cases can be exemplary of the hidden role of chance (Taleb, 2001), or broader issues of our global risk society (Beck, 1992, 2000).

There can be found in the literature another perspective defining exemplarity as signalling to higher values and civility in the context of moral education. That is the case of mythological and religious figures as well, who represent so-called moral *exemplars* as defined by Zagzebski (2013). This perspective may be included in the motivational tradition, but, as described in her *exemplarist virtue theory* (Zagzebski, 2010), the specific motivating emotion of an exemplar to students is admiration. Moreover, Korsgaard (2019) draws a line between the first three (didactical) perspectives and the fourth (pedagogical) one in educational practice. The didactical perspectives refer to some*thing* being exemplary, while in the pedagogical some*one* takes up an exemplary function. As I am looking for principles for designing tasks (some*thing*), the moral perspective is therefore not considered.

Overall, when it comes to the design of probability and statistics learning environments, the exemplary principle should be applied for choosing cases whose quality relies on balancing three orientations. An example must be relevant to the students' personal lives, as a case that can **subjectively** relate to them. The case must be **instrumental** for the disciplines involved, illustrating and connected to probabilistic and statistical big ideas. Finally, contextualised examples should be exemplary of broader societal structures, allowing for a **critical** stance towards society's shaping through probability and statistics. The latter perspective is crucial to living up to the critical justification discussed on Chapter 3. By applying Negt's perspective to general education, deliberation or decision-making in the social and political realm can only be enabled by school probability and statistics if such a practice illustrates those broader systemic structures. The first design principle allows us to choose and evaluate the quality of exemplary cases, but not yet about a concrete structure of the teaching situation that relies on exemplarity. The following principle intends to.

4.2.2 Inquiry approach

As discussed in subsection 2.3.2, inquiry-based mathematics education is my chosen framework for action. As guidance for designs, an inquiry approach comprises a **guiding investigation**, usually formulated as a question, which should be possible to approach by the disciplinary content intended to be learned.

The second aspect to the inquiry approach is that it gives **structure** to the designs, based on the three basic principles as described by Blomhøj (2013, 2016): (1) setting the scene, (2) students' investigative work, and (3) joint reflection. These principles entail:

1. Setting the scene towards the activity:
    - The general purpose of the research is shared with the students.
    - An inquiry/problem/challenge is explicitly given to the students, hopefully, posed as a question.
    - Time and space conditions are discussed.
    - General instructions about the sequence are read and clarified out loud.
    - Special attention is given to expectations: What should be returned as a product by the end of the sessions?
2. Students' independent investigative work:
    - Students are given tasks to work in groups.
    - Blank papers, calculators and other materials are available.
    - Teacher and researcher model exemplary dialogues.
    - Several concepts are ambiguous, such as risk, uncertainty, fairness. Teacher and researcher must be prepared and establish a way to engage these likely confusions.
3. Joint reflection and mathematical learning:
    - Each group of students has the chance of sharing their experiences.
    - The teacher systematises mathematical notions evoked during the sessions, connected to previously established knowledge.
    - All participants are allowed to propose new inquiries based on the experience.

(Blomhøj, 2016, p. 156, my translation and adaptation)

I draw upon these principles as guidance or point of departure for the chronological organisation of the designs as didactical sequences. However, they are not meant to be linearly applied in a strict sense, given that, for example, the systematisation of mathematical notions could be needed at different stages of the sequence, and sharing results and experiences can be relevant at the beginning of each session. For example, in the "Should we install a thermoelectric power plant?" task (Chapter 7), a consensus about the probability of a single-day environmental disaster is needed for a calculation in the long term, so a discussion in plenum takes place to clarify issues of independence and conditional probability in the calculation. This continuous scaffolding practice is what Hmelo-Silver, Duncan, and Chinn (2007) claim to be one of the main distinctions between inquiry approaches and unguided discovery learning.

Inquiry approaches to teaching and learning rely on the **authenticity** of the task situation, "in terms of connections with students' real life and link with out-of-school questions and activities" (Artigue & Blomhøj, 2013, p. 809). There are several views on the authenticity of tasks appropriate to mathematics, probability and statistics, whether it is the personal and cultural authenticity in risk situations (Murphy, Lunn, & Jones, 2006) or the use of real or realistic data in statistics learning environments (P. Cobb & McClain, 2004; Ben-Zvi, Gravemeijer, & Ainley, 2018).

A thorough discussion on authenticity can be found in Palm (2009), mainly oriented to word problems in mathematics. Palm proposes a framework for evaluating the authenticity of tasks in terms of their *representativeness* (a combination between *comprehensiveness* and *fidelity*) to the corresponding real-world task-situations. In this thesis, I focus on the fidelity aspect alone, from the standpoint of exemplarity. He provides a set of aspects where this concordance is to be assessed, namely *event, question, information/data, presentation, solution strategies, circumstances, solution requirements and purpose in the figurative context* (2009, p. 9). For the sake of this thesis, the problem's presentation and its circumstances are assumed to be unknown or unrealistic, in the school context. All four designs refer to real events and guiding questions in the out-of-school context. Information and data vary depending on the situation. For example, in the "Guilty or innocent?" (Chapter 5) and "How are PSU scores computed?" (Chapter 6) learning environments, students discuss the cases regarding data as were presented, respectively, on trial and media in the "Should we install a thermoelectric power plant?" task (Chapter 7), data are fictitious; and in "How many people attended the students' march?" (Chapter 8), students construct data from aerial pictures of the event. The purposes become part of the discussion itself. For instance, in Chapter 6 students are meant to unravel the politically framed purpose of graphical representations of statistical information. Finally, the solution strategies and requirements are subject to pragmatic curricular considerations, since in some cases, the mathematics involved is too complicated for upper secondary school students. This consideration leads to the following principle.

4.2.3 Pragmatism

The overall *problématique* addresses a teaching practice under conditions as genuine as possible. *Managing the gap* between research intentions and teaching conditions requires an acknowledgement that "instruction depends on the values of the participants; it depends on technological infrastructure; it depends on the nature of classroom discourse; it depends on practicalities such as available time" (diSessa & Cobb, 2004, p. 82). The *gap* can be represented as that between an *imagined situation* and the eventual arranged situation, which is the product of a collaboration with the teacher through *practical organisation* (Skovsmose & Borba, 2004, p. 218). These elements discussed with the teacher before and in-between the sessions are curricular framework, previous knowledge of the students, and time availability.

The official **curricular frameworks** at the time of planning and implementation[5] are in MINEDUC (2015b) for 11th grade and MINEDUC (2015a) for 12th grade. The contents that should be covered in 11th grade for the unit of data and chance are (MINEDUC, 2015b, p. 121): conditional probability; discrete random variables, their probability function and probability distribution; binomial distribution, its expected value, variance and standard deviation. Two units of data and chance are defined for 12th grade, one related to probability and another to statistics. For probability, the contents to be covered are (MINEDUC, 2015a, p. 85): continuous random variables, their probability distribution and density functions. As for statistics, the enlisted contents are (MINEDUC, 2015a, p. 115): normal distribution, Φ cumulative function, estimation of population means, and construction of confidence intervals.

The curricular aspect of pragmatism is meant as an upper bound for the knowledge base needed to address the inquiries. For example, a fascinating case is the one discussed by Fenton, Neil, and Thieme (2018), where empirical yearly average deaths caused by jihadist terrorist attacks and lawnmower accidents (among others) in the USA are compared to make the case against the so-called Muslim travel ban imposed by the federal government in 2017. The comparison of sample means goes beyond the Chilean upper secondary school curriculum. Moreover, the core of the technical debate is that comparison methods assume random variables to come from a distribution with finite variance. However, seldom violent events can more reasonably be seen as *fat-tailed* (see, e.g. Cirillo & Taleb, 2016). Given the exploratory character of this project, the consideration of the curricular framework does not mean a strict adherence to cover it, but the possibility to turn it into a didactical design for teaching practice within the curricular boundaries.

Previous knowledge of the students is related to the curricular framework but much better informed by the teachers and their particular teaching plans. Prioritisation of content and depth is a professional decision based on the complexity of factors that I refuse to judge and alter. Instead, adaptations were made. For example, both tasks oriented for 12th grade were proposed to the teachers initially counting on the students' familiarity with the normal distribution. However, according to both teachers' plans, by the time scheduled for implementation, they either had not covered it or decided not to. From there, in Chapter 6 the normalised scale is bypassed through a constructed table, and for the confidence intervals in Chapter 8, the value and meaning of the normal score $z_{0.975}$ are simply given as part of a formula. This pragmatic aspect is meant to be a lower bound in the students' knowledge base, i.e. a presumed point of departure.

Finally, and in the spirit of imagining genuine teaching practices, the tasks ought to take place during regular mathematics class schedules. We are not fooling anyone: students are aware that they are participating in a research project, but the circumstances should be as quotidian as possible. The negotiation with teachers led to an availability of one week worth of regular mathematics classes (supplementary upper secondary school courses of "Algebra and Analytical Models" and "Functions and Infinite Processes" are excluded), equivalent to 5 pedagogical hours (225 chronological hours) distributed into three slots, by the time teachers were meant to teach the "data and chance" unit of the mathematics programme. This is our **time availability**.

4.2.4 An overview of the tasks

The principles stated above inform the design of four inquiry-based learning environments, corresponding to the titles of the upcoming chapters, where they are described in further detail.

In what follows, I describe them briefly, pinpointing how they connect to the design principles.

Chapter 5: "Guilty or innocent?"

In order for a person to be declared guilty, the principle of presumption of innocence must be applied, and to *prove* her guilty, it must be beyond a reasonable doubt. This logic mimics that of hypotheses testing, where statistical evidence in favour of a claim (guilt) must be very unlikely to occur if the null hypothesis (innocence) is assumed to be true. Students are encouraged to reflect on a real case where a mathematical calculation was used in the context of a wrongful conviction. To do so, they first design an experiment to decide whether a classmate can distinguish juice with and without sugar. The intention is to discuss the transfer of the logic of hypothesis testing to justice.

By designing an experiment and collecting their own data is undoubtedly exemplary from a subjective perspective, since students can relate to their own classroom experience. At the same time, the juice experiment is instrumental for ideas of data and probabilistic evidence. However, the aim is to reflect upon a landmark judicial case; namely, the case of Malcolm Collins (People v. Collins, 1968), which is not close to the students' reality, as it took place in the United States. Moreover, prejudices of race, which would be relevant to understand the case (the suspects were an interracial couple in the 1960s), are not so evident in Chile. On the other hand, the case is well documented and exemplary for the use of the logic of hypothesis testing as probabilistic evidence, involving the issues of when to multiply, the prosecutor's fallacy, and the 'out of how many' principle (Rosenthal, 2015). These three issues echo more general probabilistic ideas than those connected to the law. Respectively, these ideas are connected to the assumption of independence (Borovcnik, 2017), the misunderstanding of P-values (White & Gorard, 2017), and the frequentist meaning of probability (Batanero, 2005; Batanero & Díaz, 2007a). As for the critical perspective, the case of Malcolm Collins is critically exemplary for the (mis)use of probabilistic evidence in justice (Gigerenzer, 2002), i.e. how probability calculations shape the judgement on a person's innocence. In other words, beyond its flaws, the case is exemplary for how the risk of wrongful conviction is mathematised.

Though the title of the task suggests the guiding investigation to be a judicial decision, the inquiry approach taken is based on two investigations. Students first investigate whether one of their group mates can distinguish between juice with and without sugar, based solely on their flavour. The idea is, once this investigation is finished, students analyse a real judicial case where a decision has already been taken. Following the general guidelines in Blomhøj (2016), the intervention begins by making this purpose transparent for the students, followed by their independent work in designing and carrying out the juice experiment. Teacher and researcher guide them to define probabilistic evidence as a P-value based on the binomial distribution. Finally, we read and discuss the case of Malcolm Collins in light of the methods and conclusions form the juice experiment. Both inquiries are authentic as referring to real events (own experiment and real judicial case) and questions (to provide evidence for a), real data (self-produced and extracted from the legal case), and solution strategies and requirements (probabilistic evidence for both).

Pragmatic considerations include, firstly, that the 11th-grade curriculum for mathematics includes, among others, conditional probability and binomial random variables, and their applications to scientific problems (MINEDUC, 2015b, pp. 120–122). Second, previous knowledge comprises Laplace's law for APT probabilities, FQT probabilities as verification of AFT, tree diagrams, notions of independence and multiplication rule. The juice experiment can be approached by a binomial model, while the case of Malcolm Collins makes use of FQT probabilities and multiplication rule. Both can help to develop a notion of conditional probability as P-value $= \mathbb{P}(\text{evidence}|H_0)$ and its misinterpretation as $\mathbb{P}(H_0|\text{evidence})$. Finally, the teacher has the availability of three sessions of 90, 90 and 45 minutes, respectively. Accordingly, the sequence is planned as:

1. Brief presentation the case ought to be reflected upon on the last session and its relation to probability. Design of an experiment to assess whether a group member can distinguish between a sugary and non-sugary juice. Collection of experimental data.

2. Computation of the P-value for the best result, under a binomial distributed null hypothesis.

3. Discussion of the case of Malcolm Collins and his wrongful conviction, connecting to the notions of independence, probability and context.

Chapter 6: "How are PSU scores computed?"

PSU is Chile's battery of standardised tests for admission to higher education institutions. Scores are computed and reported using a normalised relative scale. Every year, PSU scores take over the media and political arguments as being an illustration of Chile's educational and broader social reality. Students are invited to understand how these scores are computed and, through its technicalities, they critique related statistical messages in the form of graphical representations in the media.

PSU is certainly exemplary in the subjective sense since the 12-graders are in preparation to take the test by the end of their senior upper secondary school year. From an instrumental perspective, PSU is exemplary for the use of statistical measures of order (e.g. quantiles), distribution of data (e.g. normal) and graphical representations of data. What is critical about the case is, above all, the prescriptive role and function of mathematical models in technological societies (Davis & Hersh, 1986, p. 120). Test results are not modelled via a normal distribution as a description or prediction of a situation, but as a tool to take a course of action, in this case, the assignation of high school seniors into higher education institutions. PSU also shows how results of standardised tests shape our views on kinds of people (e.g. social class) (Popkewitz & Lindblad, 2018), how some patterns are not observed in data but designed (e.g. normally distributed), and the graphical exercise of political arithmetic (Donnelly, 1998).

The inquiry approach is looser in this case since the ultimate purpose of the sequence is not stated in the guiding question "how are PSU scores computed?", but in the interpretation of the graphical representation of PSU score trends in Figure 6.3. In hindsight, the inquiry should have been stated as "do PSU score trends imply our educational system is really hurting?" Nevertheless, the activities begin by announcing the overall goal of analysing a graph of PSU score trends and setting the scene by reading and discussing a news article containing summarised data of PSU results from the previous year. It is followed by students independent work on computing PSU scores based on their own practice tests considering different cohorts. Finally, students interpret and construct their own graphical representations of PSU trends and their political usage. Concerning authenticity, as an event, PSU is simply a fact of life and so are the summaries and graphs on the media. The questions posed in the problem are not authentic, since the computation of PSU scores is not a problem, but the application of an algorithm. However, the question of whether a particular trend represents bad news for the educational system was an authentic matter of national political debate. Data (PSU results, news articles, tables and graphs) come from official sources, though the search for further necessary data may improve the task. The solution strategy for computing PSU scores is taken from the official algorithm, but generalised from the version of 2017 of the PSU test in Mathematics (Table 6.2).

In terms of pragmatism, the 12th-grade curriculum expects as a learning goal to critically evaluate statistical information from the media (MINEDUC, 2015a, pp. 84–86) and applications of the normal distribution (MINEDUC, 2015a, pp. 115–117). Graphical representations have been an organic component of their education in different subjects, and position measures are familiar to them, as income quintiles play a role in the access to economic benefits. We count on three sessions of 90, 45 and 90 minutes, respectively. Accordingly, the sequence is organised as follows:

1. Setting the scene by reading a news article of PSU results of the previous year, sharing the general inquiry and analysing Table 6.2 to get an understanding of the percentile-based computation of PSU scores.

2. Computing PSU scores based on students' latest practice tests, according to three cohorts: their working group, their whole class, and the previous year's national cohort.

3. Evaluating critically Figure 6.3 by focusing on technicalities (e.g. axes) and its interpretation based on an understanding of PSU scores.

Chapter 7: "Should we install a thermoelectric power plant?"

A recent environmental disaster in a neighbouring province sets the scene for students to discuss the notion of decision-making under risk. A fictitious company uses its (low) risk calculations to defend the installation of a new thermoelectric power plant. Students are set to verify probabilistic calculations involved, explore their meaning and decide by themselves, illustrating the complexities beyond the notion of risk.

The inquiry is contextualised in an ongoing project for the instalment of a thermoelectric power plant in the students' province, neighbouring one that just suffered from an environmental disaster (Televisión Nacional de Chile, 2018), thus being the case exemplary from a subjective perspective. From an instrumental perspective, the case is exemplary for the use of FQT probabilities, underlying assumptions of independence, conditional probabilities, and decision making in situations of risk. From a critical perspective, the situation is exemplary for illustrating several issues of the risk society, including the distinction between risk and uncertainty situations (computable and non-computable), the dual nature of risks as both factual and value statements, and the way the past loses its power to determine thought and action(Beck, 1992, 2000).

The inquiry is stated in the title in the form of a decision and structured roughly in three stages. First, the scene is set by recalling the recent environmental saturation declared in the province of Quintero and Puchuncaví (TVN, 2018) and present in the general inquiry about the decision of a new power plant in their province. Second, based on provided data, students compute the probability for a disaster to occur on a single day and in the long term. Finally, a decision must be made. The one aspect truly authentic this time is the event and guiding question, namely an ongoing fossil-fuelled power plant in the students' province and a decision must be made. All data and solution strategies are fictitious for lack of access and adaptation to the school curriculum.

The 11th-grade curriculum covers the application of conditional probabilities and discrete distributions such as Bernoulli and binomial. Under appropriate assumptions, once computed a probability of disaster in a day, the binomial model can provide an extension to the long term. From the teacher's input, students are delayed in some of their probability backgrounds. Nevertheless, at the point of implementation, they are familiarised with Laplace's law for APT, computation of FQT probabilities and the multiplication rule. We count with three sessions of 90 minutes. However, the second session was reduced at the time of implementation, due to a misunderstanding of the use of rooms in the school. Accordingly, the sequence is organised as follows:

1. Setting the scene by recalling the disaster taking place in Quintero and Puchuncaví, and informing the new project in their province. Evaluating the probability calculation of an environmental disaster to happen on any day, as given by the fictitious company. Producing students' own calculations after appropriate corrections.

2. Expanding of the previous calculation onto a longer time span.

3. The making and grounding of the decision embedded in the general inquiry; should we install the power plant or not.

Chapter 8: "How many people attended the students' march?"

Inspired by the "Counting people" task from the Mascil Project[6] and reported by Triantafillou, Psycharis, Bakogianni, and Potari (2018), students are asked to find an estimate for the number of attendees at a students' demonstration taking place on April 2018 in Santiago, Chile.

The case is exemplary from a subjective perspective, not just because it relates to a students' march. It turns out that many of the students in the class actually attended to it. The task is instrumentally exemplary for estimation based on statistical inference (Zieffler, Garfield, Delmas, & Reading, 2008; Makar & Rubin, 2018). From a critical perspective, the case is exemplary for how crowd size estimations could shape our views on the causes for which the crowd is gathering.

The guiding inquiry is stated on the title. As for the structure, the activities begin by reading news articles referring to the students' march and different estimates of attendees to set the scene. Students' independent work is scaffolded by constructing, sampling and making inferences about the crowd size based on aerial photographs. Finally, the joint reflection encompasses sharing results and insights about the activity. The event referred to is authentic, and so is the question. As much as authorities do use aerial pictures to make estimates, which pictures and what procedures they use are not transparent to the public. The same can be said about the methods of estimation.

The 12th-grade curriculum includes, as a learning goal, the evaluation of statistical information from the media, and distribution of sample means and normal confidence intervals as contents to be covered. The approach taken is precisely to steer towards the use of confidence intervals as the estimation strategy. Students are familiar with histograms, and notions and calculations of central tendency and spread measures. We count with three sessions of 90, 90 and 45 minutes, respectively. We organise the sequence accordingly:

1. Setting the scene with news articles that show estimates for the number of attendees to the event. Estimating the number of people using a closeup picture.

2. Extrapolating previous results into a broader picture of the entire demonstration.

3. Presenting and discussing results contrasted to estimations from the scene-setting of the inquiry.

4.3 Collaborative classroom interventions

The four inquiries were implemented in close collaboration with three different teachers in Chile, from now on referred to as Alejandra, Bernardo and Carla. The contact was made quite differently in each case. I contacted Alejandra through her school's headmasters, which I knew personally from many years. Bernardo's contact was given to me by a fellow PhD student I met during one of my academic stays in Chile. Carla had approached me during a local conference, manifesting her interest in implementing some of my ideas at the time. All of them were asked to participate by presenting the general research goals and with a list of inquiry suggestions that could fit their and their students' interest and teaching plans.

At the time, Alejandra was teaching mathematics for both 11th and 12th-grade classes. With her, we developed the "Guilty or innocent?" and "How are PSU scores computed?" tasks (Chapter 5 and Chapter 6), respectively. The "Should we install a thermoelectric power plant?" (Chapter 7) task was implemented with Bernardo's 11th-grade class. Finally, "How many people attended the students' march?" (Chapter 8) was carried out with one of Carla's 12th-grade classes. All implementations took place in the period of October-November 2018, by the end of the school year, where the part of the programme related to data and chance is often taught.

During the classroom interventions, teachers had different degrees of participation, but both researcher and teacher were present in the classroom the whole time. Roughly, I as researcher took the responsibility of presenting the intention of the study and the general inquiry. The teachers dominated the whiteboards whenever we decided a pause was needed to clarify mathematical definitions and calculations to the whole class. In other words, teachers kept their role of mathematics teachers for their students.

4.3.1 Production of data

The purpose of this project is to see to what extent it is possible to evoke critical reflections by students, and dialogues supported by their work are the units of analysis. Therefore, data were collected in the form of audio and video recordings and written work by the students for all classroom sessions. One Dictaphone was positioned at the centre of each working group and a video camera recorded from a corner of the classroom, as support for later transcription. Identifying who is talking at what moment can be challenging without visual synchronising.

I make all transcriptions from the classroom interventions. Due to the volume of video and audio and the focus on critical reflections, I fully transcribed the joint reflection phase of each classroom experience, as a point of departure. From there, in the process of analysis, I draw on video, audio and field notes of previous episodes when needed.

4.3.2 Ethical considerations

Teachers have a duty to cover a curriculum and students are entitled to learn the content they are supposed to in the time available. The exploratory character of my research and inquiry-based approach to the learning environments can compromise that expectation. In particular, the classroom experiences take place in the context of 11th, and 12th-grade classes and access to higher education in Chile is granted by PSU (the standardised test we address in Chapter 6), which students take at the end of their upper secondary school year. Spending mathematics classroom time participating in this research instead of preparing for exams may be harmful to their results. Regardless of my aim to try out a teaching practice with a potential that goes beyond succeeding on a test, in the short term, these exams are currently the means for students to accomplish their goals. I address this issue employing the principle of pragmatism described in Section 4.2 and by making my intentions very explicit at the moment of proposing the teachers to participate in the project.

In terms of transparency and privacy, informed consents were signed by all three teachers and students' parents or legal tutors. Students who were already 18 years old at the time of the interventions signed the consents themselves. They signed their consents themselves. In these signed documents, participants accept full and freely their participation in the PhD project, being informed of the following:

- Research is being made with the exclusive end to promote scientific knowledge. Teachers count with a proposal, including theoretical and practical considerations.
- Data are collected in the form of audio and video during involved class sessions.
- Audio and video will be watched, listened to and process solely by the researcher and eventual close colleagues.
- Names will be kept confidential, and any information that could identify them will not be presented nor published. Instead, pseudonyms will be used for all type of dissemination.
- They have the right to decline the use of collected data from them at any moment and without the need to provide any reason.
- They are given the contact information of the researcher for any kind of consultation.

Additionally, they authorise the use of data for future publications after the PhD project, under the same conditions of confidentiality.

4.4 The teachers' input

Since critical research possibilities are facilitated by "researching *with*, and not *on* teachers" (Skovsmose & Borba, 2004, p. 220, emphasis in original), I carried out post-intervention interviews with the teachers about the entire process, positioning the teacher as a 'research subject' as opposed to 'research object'. The general aim of this stage of the research project is to get the teachers' insight into the classroom interventions that took place. It includes the micro-level interpretation of students' work, evaluation of the task itself, the collaborative work with the researcher, impact of the experience on students and teaching practice, and the potential of projecting elements of the experience into the future.

In order to collect the teachers' honest and spontaneous views, whilst keeping a focus in particular research topics, semi-structured interviews were the choice for a method. A semi-structured interview, together with unstructured interviews, lies within the category of qualitative interviews. Bryman defines it as follows:

> The researcher has a list of questions or fairly specific topics to be covered, often referred to as an *interview guide*, but the interviewee has a great deal of leeway in how to reply. Questions may not be asked exactly in the way outlined on the schedule. Questions that are not included in the guide may be asked as the interviewer picks up on interviewees' replies. But, by and large, all the questions will be asked, and a similar wording will be used from interviewee to interviewee.
>
> (Bryman, 2015, p. 468, emphasis in original)

Accordingly, the next step was to construct an interview guide. Following the review by Kallio, Pietilä, Johnson, and Kangasniemi (2016), a semi-structured interview guide should be organised in main themes in a progressive, logical order. In this case, the main themes are (1) interpretation of students' work, (2) potential and possibilities; and (3) collaboration between researcher and teacher; in that order. The first theme is of particular relevance both as an *aide-mémoire* and for the analysis of students' work, as will be explained on Section 4.5. The second theme is meant to account for the teachers' perspective on projecting the classroom experience into the future. The third theme regards the circumstances and challenges of participating in the study itself.

Within each theme, sub-themes are established as well, and follow-up questions fulfil the role of clarifying the main themes and gaining accurate and optimal information. For the sake of valuable and honest information, Bryman (2015) claims that leading questions should be avoided. So instead of asking directly, one should have questions that prompt desired types of answers. Direct questions often trigger biased and poor data. On the other hand, they can be useful to set the scene and start the discussion. Regardless, interviewees are aware that they are expected to provide richer answers. For each sub-theme, I have a **question** as if it was of the direct type, and a **prompt** to be used in case the information from the interviewee does not come spontaneously. The full interview guides with themes, sub-themes, questions and prompts are displayed in Appendix A. During each interview, I used a printed copy of the interview guide to make time marks next to the sub-themes as the interviewees' remarks matched them.

Two teachers accepted to participate in the interviews, namely Bernardo and Carla. Their interviews took place during December 2019, more than one year after. Each interview had a duration of approximately two hours. Alejandra declined the interview on the grounds of being on her maternity leave at the prospected time of the interview. After clarifying that it does not need to be at that precise time nor her workplace, she did not follow up. My personal impression is that she was not interested in participating anymore. A possible explanation can be the fact that, as opposed to Bernardo and Carla, her participation in the project was asked not directly but through her headmasters. Therefore, her involvement in my research came across as her bosses' request instead of personal and professional interest.

4.5 Analysis

In Chapter 2, I have pointed out that addressing **RQ4** implies detecting whether students show signs of reflecting upon the formatting power of mathematics. In the context of mathematics education, a *reflection* can be understood as "a deliberate act of thinking about some actual or potential action aiming at understanding or improving the action. Reflections take place in the minds of individuals but are strongly influenced by social interactions, and *they can only be detected and analysed through communicative acts*" (Blomhøj & Kjeldsen, 2011, p. 386, emphasis added). I highlight the argument that the units of analysis are participants' (mostly students') discussions. In order to interpret these discussions as evidence for reflective knowing in different levels, I borrow the six entry points introduced by Skovsmose (1992, 1994, 1998) and discussed in subsection 2.3.3. Proposals for their operationalisation have been suggested by, e.g. Barbosa (2006) in the context of mathematical modelling in general and Hauge, Gøtze, Hansen, and Steffensen (2017); Gøtze (2019) in socio-scientific discussions about climate change. In all these cases, knowing can be of the mathematical, technological or reflective type, depending on whether reflections are mathematics-, model- or context-oriented (Skovsmose, 1998). These orientations correspond, respectively, to entry points 1 and 2, 3 and 4, and 5 and 6. According to Skovsmose (1998, p. 199), reflections are considered to be critical when the scope of reflections is oriented to "political, social or cultural context". Accordingly, my claims of critique taking place in the classroom experiences are evidenced by students' context-oriented reflections, henceforth **critical reflections**, i.e. when the subject of reflections addresses matters of the formatting power of mathematics in society. For example, in Chapter 7, mathematical reflections are elicited by students' checking results of calculations of auxiliary probabilities of an environmental disaster. Technological reflections are elicited by discussing the assumption of independence of events leading to multiplying their respective probabilities. Critical reflections are elicited by students' realising that assuming independence of events is convenient for the company, as it affects the perception of risk of an environmental disaster.

However, as stated in **RQ5**, my goal is to confront the critical justification of probability and statistics in particular. Therefore, a further layer of the analysis is made to pay exclusive attention to issues of probability and statistics. In the planned exploratory spirit of this project and unplanned reality of ongoing learning as it is carried out, I draw elements of probability and statistics in different ways for each of the classroom experiences, which I made explicit in subsection 2.3.3.

In Chapter 5, I identify the three main issues of hypothesis testing in the context of the risk of wrongful conviction as defined in Rosenthal (2015), which inspired the didactical sequence in the first place. These concerns are the ‘when to multiply?’ question, the ‘prosecutor’s fallacy’ and the ‘out of how many’ principle, which are specific interpretations of the problem of the assumption of independence (e.g. Borovcnik, 2015), the misinterpretation of P-values (White & Gorard, 2017) and the FQT meaning of probability (Batanero & Díaz, 2007a), respectively.

In Chapter 6, I do not draw on any particular framework defined *a priori*, but instead, I draw on critical and historical aspects of statistics to make sense of students discussions. In the joint reflection regarding the use of a graphical representation on the media, we even develop a local proto-framework for analysing and constructing graphs as the activity goes. We achieve this framing for the analysis in three steps. First, as a researcher, I hold Skovsmose’s entry points (1992, 1994) on reflective knowing as guidance for choosing exemplary graphs. In preparation of the task with Alejandra, I am explicit about looking for introductory graphs that can elicit these three levels of discussions: mathematical-oriented, model-oriented and context-oriented. Respectively, there should be introductory examples where there are issues of right or wrong depictions, where political messages can be inferred, and where the usage of mathematical tools can be questioned. Second, during the classroom experience, with these issues in mind, both teacher and researcher guide the conversation accordingly, letting students find what is right or wrong in the graphs, and the intentions therein. We take notes on the whiteboard after students make observations, i.e. cherry-picking years on the X-axis or what politician is hurt by the message given. At the moment of analysing Figure 6.3, we remind students of their considerations given in the introductory graphs. Finally, after the transcription, I summarise and make use of this framework to analyse students’ discussions, connecting its categories to the notion of reflective knowing (Table 6.4).

In the next two Chapters, I do a more systematic process of coding based on *a priori* frameworks. I also make use of Bernardo and Carla’s interviews for further sense-making of the students’ work. In some cases, we review interesting episodes that allow me to *triangulate* the coding process and enrich the discussion of results. This process is one humble way of improving the reliability in the analysis, without having teachers to participate in the research project as coders. That would be abusive and worth my emails unanswered. Scotland (2012, p. 13) argues that this is a common imbalance in interpretive research, since “though participants are often given a voice, it is usually the researcher who decides on the direction that the research takes, the final interpretation of the data, and which information is made public”. In Chapter 7 I code in relation to the aspects of probabilistic thinking and probabilistic literacy in the context of risk as defined by Borovcnik (2017). Finally, in Chapter 8 I attempt to operationalise the critical statistical literacy framework as delineated by Weiland (2017), broadening the scope of literacy from reading the world with statistics to writing it.

For the writing of the thesis, all translations from Spanish are my own and made preserving its function in the communicative context. Initial efforts to communicate my research with a close-to-literal translation of excerpts did not succeed to be understood. It should not be surprising, considering that it had a similar effect on Spanish speakers from countries other than Chile, without translation.

In summary, this thesis is motivated by Critical Mathematics Education, a research programme that draws on a critical paradigm to identify its preoccupations. However, this research project is rooted in an interpretivist research paradigm, using a qualitative strategy, using a methodological design of a multiple-case study. Each of the four classroom experiences defines a case.

I define three principles; exemplarity, inquiry approach and pragmatism. These guide the design of learning environments, aiming to live up the critical justification for probability and statistics in the Chilean upper secondary school context, thus playing an initial role to produce the data to address **RQ4** and **RQ5**. These principles will also be revisited in Chapter 9 to address **RQ3**.

The methods consist of the implementation of four classroom interventions, based on the principles aforementioned, and designed in collaboration with three teachers, Alejandra, Bernardo and Carla. The four learning classroom experiences are named after their guiding inquiries, namely "Guilty or innocent?" (Chapter 5), "How are PSU scores computed?" (Chapter 6), "Should we install a thermoelectric power plant?" (Chapter 7) and "How many people attended the students' march?" (Chapter 8).

All three teachers and their pupils' tutors signed informed consents to participate in the research project and accepted the use of anonymised video and audio data. In the case of Bernardo and Carla, they accepted to take part in semi-structured interviews, to enrich the analysis and have their insight into projecting a similar practice in the future.

In order to address **RQ4** and **RQ5**, the analysis is made by coding in two layers. First, Skovsmose's (1992, 1994) entry points on reflective knowing specified in subsection 2.3.3. Second, to address the specificity of probability and statistics in the coherence along the *problématique*, I make use of different domain-specific frameworks, also unfolded in subsection 2.3.3.

Notes

[3]This research used the databases of the Agency for Quality in Education (Agencia de Calidad de la Educación, 2016, 2017) as a source. The author thanks to the Agency of Quality in Education the access to information. All results of the study are the author's responsibility and do not compromise the Institution in any aspect.

[4]Source: Agencia de Calidad de la Educación (2019).

[5]A new curricular framework has begun to rule from the school year of 2020 (MINEDUC, 2019).

[6]Available in http://www.mascil-project.eu

Chapter 5

“Guilty or innocent?”

> Mathematics, a veritable sorcerer in our computerised society, while assisting the trier of fact in the search for truth, must not cast a spell over him. We conclude that on the record before us defendant should not have had his guilt determined by the odds and that he is entitled to a new trial. We reverse the judgement.
>
> (People v. Collins, 1968)

Statistical hypothesis testing is a standard quantitative approach to convey scientific evidence. If, for example, one is testing for the effect of a drug, it is not enough to show that it works, on average, better than the placebo. It is reasonable to ask whether the difference D found between the treatment and control subjects of a clinical trial was due merely to sampling error. That is, even if we assume the drug and the placebo to have the same effect in the population (the null hypothesis H_0), we expect them to show different results when taking two random samples. We need to go further and ask how likely it is to obtain a result at least as *convenient* as D, assuming that the null hypothesis H_0 is true. If this probability $\mathbb{P}(D|H_0)$, called P-value, is small enough, then data are claimed not to be compatible with H_0, and thus it can be rejected in favour of the alternative, i.e. the drug has a different effect than the placebo with statistical significance. If the P-value is not small enough, data cannot rule out H_0.

This rationale is analogous to the legal principles of the presumption of innocence, demanding to convince a jury of the opposite “beyond all reasonable doubt” (Greer & Mukhopadhyay, 2005; Rosenthal, 2015). When a suspect is put to trial, in order to declare him or her to be guilty, it must be proven beyond a reasonable doubt, i.e. the gathered evidence should be contrasted with the presumption of innocence and make the case that it is very unlikely or unreasonable for the evidence to be compatible with the suspect’s innocence. It is no surprise that connections can be drawn between probability and the law. For example, top mathematicians such as Condorcet, Laplace and Poisson had a significant influence in the configuration of the French jury during the 19th century, by computing the probabilities of errors in conviction decisions as a function of the divide in the jury (Hacking, 1990, Chapter 11). Part of the classic development of probability was in the context of the law.

Rosenthal (2015) shows how the problematic use of mathematical calculations has led to actual wrongful convictions, by highlighting three issues: the “when to multiply?” question, the prosecutor’s fallacy and the “out of how many” principle. These problems have a more general typification in probabilistic terms (e.g. Borovcnik, 2015, 2017). The multiplication question is embedded in the assumption of independence of events. Pieces of circumstantial evidence can be arbitrarily added, and their probabilities multiplied as if independent, in order to convey a small probability when convenient. Such is the case of a prosecutor attempting to prove a suspect to be guilty. The prosecutor's fallacy (Thompson & Schumann, 1987) is a misunderstanding of the probability that yields the circumstantial evidence gathered under the presumption of innocence $\mathbb{P}(\text{evidence}|\text{innocence})$ as if it were the probability of the suspect to be innocent, given the evidence gathered $\mathbb{P}(\text{innocence}|\text{evidence})$. The former is the information available; the latter is the information we would need. In hypothesis testing, the P-value, defined as $\mathbb{P}(D|H_0)$ is widely misunderstood – even among experts (Erickson, 2006) – as $\mathbb{P}(H_0|D)$, which are not interchangeable nor can one deduce one from another (see, e.g. White & Gorard, 2017; Gorard & White, 2017). Finally, even if the P-value is well calculated and interpreted, its value must be interpreted in terms of the population. Out of context, a P-value of, say, 0.001 may appear to be strong evidence, but in the context of a big city, the likelihood of at least one innocent person to match the gathered circumstantial evidence being 1 out of 1,000 people is not convincing. The "out of how many" principle can be thought in terms of a lottery. The probability of a particular person to win may be small, but the probability for someone to win is not.

These problems provide an opportunity to invite Alejandra's 11th-grade class to critique the use of this logic of probabilistic evidence in a scenario of the risk of wrongful conviction, being critically exemplary for how probability applications shape the perception of guilt or innocence of suspects. Albeit authentic and critically exemplary, a judicial case may not be of the students' close interest. A more subjectively exemplary problem should be addressed before analysing the judicial case.

A further tension is met between instrumental *exemplarity* and *pragmatism*. On the one hand, addressing the issues presented above is certainly exemplary for disciplinary methods and concepts, namely conditional probabilities, hypothesis testing, and P-values, among others. On the other hand, hypothesis testing is not part of the curriculum, let alone the notion of P-value; while independence, conditional probability and the binomial distribution are. Moreover, we expect that three sessions in one week would not be sufficient to cover these matters thoroughly and applying them for the judicial case study. Our agreement and our challenge are not to be necessarily explicit about hypothesis testing in their formalities, but to scaffold the logic of hypothesis testing and P-values through a particular case.

With these elements in mind, we decide to let students design and analyse an experiment relying on a binomial distribution for testing hypothesis, i.e. an experiment of repeated independent trials with binary outcomes. After some deliberation, we decide for testing whether a student can distinguish between sugary and non-sugary juice of the same brand. From that point, we can make connections to the judicial case. Thus, our arranged situation (Skovsmose & Borba, 2004) is turned into a didactical sequence into three available sessions of 90, 90 and 45 minutes:

1. Brief presentation the case ought to be reflected upon on the last session and its relation to probability. Design of an experiment to assess whether a group member can distinguish between a sugary and non-sugary juice. Collection of experimental data.

2. Computation of the P-value for the best result, under a binomial distributed null hypothesis.

3. Discussion of the case of Malcolm Collins and his wrongful conviction, connecting to the notions of independence, probability and context.

5.1 Session 1: Numbers do not speak for themselves: What counts as data

It is widely accepted that the notion of data is inextricably connected to statistical reasoning, to the lengths of portraying the activity of statistics as learning from data (Moore, 2010). Batanero et al. (2016) observe that in probabilistic reasoning, it may be counter-intuitive to associate data with chance models, i.e. the given (*datum* in Latin) with the unforeseen. One exit to this apparent tension is the so-called *structural equation of data* (Borovcnik, 2006), by which data in probability are portrayed as a composition of signal + noise, or explained + unexplained variation, which is the essence of what we do later on Session 2.

Following the design principle of *pragmatism*, Alejandra plans to cover the binomial distribution during this week. However, the pedagogical challenge is to adhere to the inquiry approach principle, to have students design an experiment to construct data in concordance with the assumption behind a binomial model, i.e. multiple independent Bernoulli-type trials. The key lies in anchoring to the context where data are collected, since "in mathematics, context obscures structure. (...) In data analysis, context provides meaning" (G. W. Cobb & Moore, 1997, p. 803). We set out to discover the binomial model, by building up from students intuitions and own work, in line with the inquiry approach.

After introducing myself to the students, I set the scene by presenting the overall idea of these three sessions briefly; to analyse a real judicial case where probabilities were involved in deciding whether a person is guilty or innocent. This prospect is not received with amusement. Especially in contrast to the expectations implanted by the presence of a video camera, audio recorders and some bottles of juice. We make clear that we need some previous steps to get into that discussion and to develop those probabilistic ideas in a more familiar context. Based on some conversations with Alejandra and the headmasters of the school, I mention the fact that a campaign of eating healthier was taking place in the school, which included the choice for products with lower sugar content:

RAIMUNDO: But many of us – and I include myself – complain about non-sugary products to be not as tasty. [*all nod and chuckle*] So, is that true?

We present the students with the available material: one bottle of juice with sugar and another with sweetener, both of the same brand and flavour. A large pile of small plastic cups are available as well. It is a small class of eleven students, and they are divided into three working groups:

- Group A: Valeria, Catalina, Nahuel, Fernanda and Marta.

- Group B: Diego, Pablo and Carlos.

- Group C: Ayelén, Gabriela and Rayén.

However, before they do anything with the material, the task is to design an experiment to test the one person from the group's ability. After nearly 35 minutes of discussion within the groups and in the plenum, an agreement for the experiment design is reached, based on three central ideas that I can make sense as: isolating the one variable at test, independence of trials, and repetition of trials. These aspects are in line with the assumptions behind a binomially distributed variable. However, during Session 1, no mentions to probability models are made yet.

5.1.1 That would be cheating

At the start, students spend some time picking who will be the guinea pig of each group. For sure, most want to volunteer to be the one drinking juice during mathematics class. So it turns into a competition over who thinks he or she can taste the difference.

Students in group C do not seem very excited, however. They ask each other "what should we do?" repeatedly until Gabriela gets in contact:

GABRIELA: [*To* AYELÉN] Do you find them both [juices] to look the same?

AYELEN: Well, I see their labels to be different.

GABRIELA: No, but their colour…

AYELEN: No. The other one has a darker colour. It can be. In my opinion, one has a darker colour.

RAYEN: [*Laughs*] Let me see... Yes! It is darker.

GABRIELA: It is clearer with sugar.

AYELEN: [Unintelligible].

GABRIELA: But that would be cheating.

ALEJANDRA: [*Walking by, to* RAIMUNDO] They say there is another factor that differentiates them: colour.

AYELEN: With a Pepsi it would have been better…

Here are some supporting dialogical features (Alrø & Skovsmose, 2002) taking place. Students get in contact, starting by Gabriela' inquiry question, which Ayelén and Rayén respond as a sign of mutual confirmation. They agree that the juice colour is different between the two types. Then Gabriela thinks aloud and advocates against an experiment where the tasting subject can tell colours apart: "that would be cheating". As a response, Ayelén is thinking aloud by proposing that differentiating Pepsi with Coke would have been "better" for the experiment's sake. I interpret this episode as locating the issue of isolating the variable at stake. It has been made clear that the experiment is exclusively about the taste of the juice, and other variables could influence the result. However, a solution is proposed:

RAIMUNDO: [*In plenum, to* GABRIELA] You say it would be cheating. So how could we solve that?

PABLO: By blindfolding them! [*general laugh follows*].

RAIMUNDO: Come again?

DIEGO: Blindfolding.

Following on with dialogical features (Alrø & Skovsmose, 2002), I take the issue of cheating as a turning point and decide to bring it to the plenum and ask an inquiry question: "how could we solve that?" Pablo proposes to use a blindfold, which is followed by a general laugh. In order to get in contact again, I support them by asking again, so Diego repeats the proposition as mutual confirmation.

In hindsight, students are building an intuition upon the validity of their experiment, i.e. the claim that we are actually measuring what we say we are measuring. In this case, we need to make sure that the ability to taste the difference is the only one put to the test. The solution of using a blindfold to isolate the taste variable by discarding the colour variable goes in line with the use of blind experiments in medical research to improve validity (Samuels, Witmer, & Schaffner, 2016).

5.1.2 Repetition conveys reliability

In order to move on with the experiment design, the leading question is about the degree of belief. Belief is connected to the subjective meaning of probability, and Bayesian inference (Batanero & Díaz, 2007b), but it is also one of the earlier ideas about probability, conceived as distinctive to opinion (Hacking, 1975, Chapter 3). Before the 17th century, a claim was said to be *probable* as in *worth of approbation*, though it used to be conveyed by ecclesiastic hierarchy. In the context of the task, degree of belief is posed as the modern concept of reliability, i.e. the claim that the result would be consistent if replicated, and some students find that a single trial is a problem. In Group A, they pick who is the test subject, Nahuel:

RAIMUNDO: How are we going to believe him?
NAHUEL: With your trust.
VALERIA: No way! It has to be with exact data.
RAIMUNDO: [*Approaching the group*] So, who is going to taste?
VALERIA: NAHUEL.
RAIMUNDO: And how is the experiment going to be?
VALERIA: He has to taste from one cup and then form the other.
NAHUEL: Then I have to try again [*group laugh*].
VALERIA: Of course! It has to be twice.
CATALINA: Twice.
RAIMUNDO: Why twice?
VALERIA: To be quite sure.
RAIMUNDO: To be quite sure... of what?
VALERIA: That one has sugar, and the other does not. He will then be confused.

I begin by challenging them with a hypothetical question, "would we believe him if he is correct?" Nahuel and Valeria begin to advocate for belief conveyed by "trust" or by "exact data". The experiment must address the results' reliability. A realisation of the experiment is given: he ought to taste two cups of juice: one with sugar, another without, and give his appraisal. Nahuel wants to keep drinking juice, and the group supports the idea suggesting that he has to do it twice. By identifying the idea of repetition, I ask a why-question, which they respond with "to be quite sure". This idea of repetition is later refined when the possibility of guessing by chance is brought up, which I discuss in the following subsection.

PABLO: I think the entire group should try [*laughs*].

RAIMUNDO: (...) So it has to be the same person. And as you said, if we give him to try once, he can merely guess at the first trial.

ALEJANDRA: So how could we know he can really distinguish between the two?

RAIMUNDO: What could we do to really believe him?

DIEGO: To try several times in different cups.

VALERIA: And to make a table with how many times he is tasting and how many of those he guesses right or wrong.

By Pablo's tone of voice, it seems he is merely fancying the idea of drinking juice. However, he does have a point, implicitly. The problem was contextualised as whether the juice itself tastes different depending on its sugar content. It was not as an evaluation of the person who tastes. A possible experimental design could be that each one in the group (or the class) gets to taste once in a survey-type of investigation. With this, the taste property of the juices would be assessed by the proportion of people who able to distinguish them correctly on a first trial ($\hat{p}$ in the sample). Appropriate methods rely on statistical inference about the population proportion p based on the normal distribution. From a pragmatic point of view, this would be a deviation from the curriculum coverage, as we intend to uncover the binomial distribution, and they have not studied continuous random variables yet. For this reason, we steer the conversation towards the question of belief in a single individual. Our pragmatic design principle obliges us to find a situation that can be modelled using the binomial distribution. The other reason for it is the connection to the judicial case we want to discuss at the end, wherein probabilistic evidence puts to test a single subject's claim for his innocence, i.e. the design principle of *critical exemplarity*.

Alejandra and I keep fishing for more trials, now in the plenum. As a result, Diego suggests the idea, and Valeria realises that the experience will require to register the number of successes and failures systematically. Afterwards, students make an arbitrary choice has been made to set the number of trials to ten. There is no further explanation other than the intuition that "the more, the better". We agree that it fits the time limitations of the session. A similar issue is risen in the "Terrible small numbers" project (Alrø & Skovsmose, 2002; Alrø, Blomhøj, Bødtkjer, & Skovsmose, 2006). In that project, students try out different sizes of egg samples to check for salmonella. In their study, the connection between reliability and the sample size is mediated by the hypergeometric probability model, which is not available to the students, so they rely on simulations that, they say "leaves them in the dark" (Alrø & Skovsmose, 2002, p. 210). In our learning environment, on the other hand, the binomial model is developed with the students during the following session. For our project, at this point, it is clear that repetition is vital to convey reliability and ten trials are agreed upon arbitrarily. Now we need to agree on how to perform each of the trials.

5.1.3 Independent trials

From a mathematical probability point of view, two events A and B are independent if and only if the probability of their intersection is the product of both probabilities, i.e. $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Since, in the general case, $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$, an equivalent definition for the independence of A and B is that $\mathbb{P}(A|B) = \mathbb{P}(A)$. That is, the probability of occurrence of A is not *associated* with or *affected* by B, among other interpretations (Borovcnik, 2017).

In order to use the binomial distribution as a model to contrast the results with, trials should be independent of each other in the real-world situation. This aspect is not straight forward to assess in the situation and "a key assumption of independence is not justified in many exercises set in probability. Indeed the key question in any modelling is the extent to which underlying assumptions are or are not justified" (Borovcnik & Kapadia, 2011, p. 2). With Alejandra, our pedagogical challenge is then to have students design their experiment justifying the assumption of independence of trials, without knowing yet that it is required for the binomial model.

Some ideas of independence had already been suggested when, for example, in group A they were proposing to do only one trial, tasting one of each type of juice:

RAIMUNDO: If he tries one and he says "no sugar", for example, and we tell him he is right, he will know the other one has sugar (. . .)

NAHUEL: The hypothalamus. . . the dopamine.

VALERIA: That has nothing to do!

NAHUEL: It does!

VALERIA: That is about drugs.

NAHUEL: Those are neurotransmitters. They communicate.

VALERIA: Haha you, the communicator! [*laughs*]. NAHUEL, we are talking about juice! And mathematics, not biology!

RAIMUNDO: [*Joins in*] Why biology? Why did you mention biology?

NAHUEL: Because [*pauses*] neurons communicate.

VALERIA: We were just talking about that in biology class.

RAIMUNDO: OK, but that is interesting. What could happen there?

. . .

NAHUEL: It could be that in the past we had already done this, so now we know how to distinguish them.

They had just been in biology class right before our session and learnt about how dopamine is behind the idea of positive reward. Nahuel realises this is somehow a problem for the sake of the experiment, locating the issue of one trial having an incidence on another. After the idea of repetition was settled, I ask them again how could the experiment go, trying to prompt a mention to the independence of trials:

RAIMUNDO: And how does it go?

VALERIA: He will try once.

NAHUEL: More than once, we said.

VALERIA: But one trial first, and you say whether it has sugar or not. Then you try again and say whether it has sugar or not. And again. . .

RAIMUNDO: You keep writing down.

VALERIA: Yes.

RAIMUNDO: Each time he is correct, will you tell him?

VALERIA: Of course not!

RAIMUNDO: Why?

NAHUEL: Because I would familiarise with the flavour.

RAIMUNDO: So trials should be...? What word can we use?

The word “independence” does not come up, even though Alejandra had been insinuating that, should they tell the student whether she guessed correctly, a “dependency” would take place. For our purposes, it is not necessary to crystallise the notion of independence at this stage, since the binomial distribution and its assumptions are not mentioned yet.

One more element arises as they begin to gather material. The designated taster will have ten trials, but if he or she knows how many cups are filled with and without sugar, then his or her attempts may be influenced by his or her previous attempts. In group A, Valeria receives the pile of cups and is ready to fill them up:

VALERIA: Alright. How many?
FERNANDA: As many as you want, but make sure to fill up ten cups.
VALERIA: No. This cannot be (...). He will then know!
NAHUEL: Let them be seven, then.
VALERIA: No, you would still know.
FERNANDA: There will be more [than five] for one [type of juice] and the other as well.
VALERIA: So he will not know if there will be more of one or the other.
RAIMUNDO: So he should know that it is possible that all [the cups he tries] could be with sugar or all could be without it.
VALERIA: Yes, that too (...) [*takes out a bunch of cups*] All of them?
FERNANDA: That will do. We can keep filling them up in any case.

In a way, Valeria thinks aloud and locates the issue of independence. Even if Nahuel does not know for sure whether he guessed correctly, he knows, in principle, that five cups contain juice with sugar and the other five contain non-sugary juice. This record could affect his judgement from one trial to another. He suggests having more than five of each. Valeria and Fernanda mutually confirm that there should be more. I paraphrase their worries as the necessity for Nahuel not to be able to assess the sugar content of juice by counting. Fernanda settles for simply filling the cups as they go, so Nahuel does not know what type of juice they are filling the cups with each time.

At this point, the experiment is designed in agreement. In each group, the selected student will wear a blindfold and be put to the test on ten trials of tasting. Each trial consists of tasting one cup of juice with sugar and another without, and decide which one is which. Their group mates should write down successes and failures. Their independent work follows, and the collection of data is done smoothly in a friendly environment. In group A, Nahuel guessed correctly on five occasions. In group B, Diego guessed eight times. In group C, Ayelén guessed three.

As the mantra of statistics education says “data are not just numbers, they are numbers with a context” (G. W. Cobb & Moore, 1997, p. 801). If data were to be taken as simple numbers in this experience, it would suffice to say that the student who guessed correctly at the most, Diego, did eight out of ten. However, the process of discussing and designing the experiment led them to take a series of considerations related to the trustworthiness of the outcomes, that go beyond numbers. Diego’s result is not simply eight out of ten successes. In context, there were eight successful guesses from the same person, out of ten independent trials, consisting of the blindfolded distinction between two types of juice, with and without sugar. As teacher and researcher, these features pave the way for introducing the binomial distribution, as it requires

- a fixed number of trials (here, $n = 10$);
- an assumption of independence between trials, conveyed by not knowing if the guess was successful each time and the impossibility of counting; and
- isolating the one binary variable at stake, the subject's taste, by blindfolding them.

Several researchers have called for the use of data to introduce probability through data (e.g. Shaughnessy, 2003). For P. Cobb (1999), using data that students can relate to can be an engaging way to ground probabilistic concepts in their particular interests and purposes. Nilsson (2014) confirms this idea by highlighting the role of experimenting, as a way to contrast theoretical and empirical probabilities. A binomially distributed random variable is still missing the assumption that each trial consists of a Bernoulli experiment with the same probability of success p. That is not what the experiment is about. Instead, it is the question we intend to ask: do data provide evidence that the student did not indeed guess by chance?

5.2 Session 2: What are the odds? Notions of probabilistic evidence

The first documented hypothesis test was published by Aburthnot (1710), in an attempt to tell whether the ratio of births between men and women can be argued to be a result of divine design or a product of chance alone. He formulates this logic by using probabilistic models to distinguish "art versus chance" (Hacking, 1975, Chapter 18) as a mathematical construct for evidence. According to the structural equation (Borovcnik, 2006), the value of using data in probability education lies in its composition of "signal versus noise", "explained versus unexplained variations", and in the ability to distinguish between them.

Batanero and Borovcnik (2016) propose that, amongst the components of probabilistic thinking that build on probability literacy, the construct of probabilistic evidence should be dealt with in high school education. They focus mostly on correlation, since hypothesis testing, as well as statistical inference in general, tends to be part of the curriculum starting in undergraduate programmes. However, the notion of *informal inference* (Zieffler et al., 2008; Makar & Rubin, 2009) encompasses a variety of problems which follow the same reasoning of statistical inference. One of these types of task is the judgement between two competing models (Zieffler et al., 2008). For example, Lee, Angotti, and Tarr (2010) use a probability simulation tool to assess the fairness of dice with students of ages 11-12, in what they define as informal hypothesis testing. Their approach is to compare the expected results of rolling fair dice against observed outcomes.

With the pragmatic principle in mind, Alejandra's 11th-grade class should be able to develop and use the binomial probability model, as well as working with conditional probabilities. Our intention is then, without a thorough definition of hypothesis testing, that students can build on probabilistic evidence through informal inference reasoning, with a taste of formality. They should be able to tell the art of correctly tasting juice, from the chance conveyed by a binomial experiment.

5.2.1 Fifty fifty

For some of the decisions made during the experiment design, the idea of telling "art from chance" already plays a significant role. For example, here is an exchange that happened right before the idea of multiple trials came to be, as a solution to the single-trial problem:

RAIMUNDO: So what is the problem? What could happen?

VALERIA: He could drink and say it is sugared and it does have sugar, but he says it just for saying it. Not because he really knew it does have sugar.

RAIMUNDO: And how likely is it for him to say that just for saying it?

NAHUEL: Fifty.

PABLO: Fifty.

VALERIA: Fifty.

RAIMUNDO: Fifty what?

VALERIA: Fifty percent.

NAHUEL: Fifty-fifty.

RAIMUNDO: So we have a problem there because it could be the case that he does (guess) it by chance. And it could be the case. So how could we solve that? If he tries once, it could still be the case that he guessed it without displaying any ability. What could we do to see whether we can truly believe him?

Valeria follows my challenge thinks aloud in a hypothetical scenario. She finds unsettling that her mate Nahuel could try once and say the right guess "just for saying it". As hard as it is to translate from informal Spanish, in context, the expression means "to muddle through" or "to get by", since Nahuel has to give some answer at some point. She contrasts it to him responding because he "knows" it has sugar. In principle, we do not know for a fact whether the person can tell the type of juice or not, but I interpret she refers to him *honestly believing* he can. This episode led to the idea of repetition discussed above.

I follow on the opportunity to introduce some language of probability, asking directly about the likelihood of such an answer to be correct if given by chance. 50% chance appears immediate, unanimous and correctly in this case. A 50% of correctly guessing is so far a criterion for not believing the testing subject. As students decided to raise the number of trials, we can build from the suggestion of having three, by challenging them:

RAIMUNDO: Here [*pointing to group C*] they proposed to do it three times. For the sake of argument, let us stay with three. And the question is: what has to happen for us to believe her (AYELÉN)?

DIEGO: It depends on how many times she was right.

RAIMUNDO: In what case we believe her and in which case we do not?

RAYEN: If there are two correct and one wrong...

DIEGO: If she is correct, let us say, three and three – three rights and three wrongs – it would be fifty-fifty. We do not believe her much.

So far, a few things are agreed. As a response to my hypothetical question, Diego locates the criterion for "believing" that a person can genuinely guess sugary from non-sugary juice depends on the number of successes. From Rayén's response to my follow-up, demanding the person to guess all trials correctly is not necessary. Then Diego doubles down on the criterion for a single trial, i.e. having half of the guesses correct is not strong evidence. He struggles with the possibility of an odd number of trials (three) because half guesses are not defined, reformulating the situation to an even number instead (six).

An intuition is built about evaluating results against an equiprobable alternative, i.e. the result of an experiment should be at least more favourable than a coin toss. At this point, students are not using a chance-related language, but referring to proportions, such as "two out of three" or "fifty-fifty". Since the experiment was designed and data are collected for ten trials, then we need to crystallise the probabilistic nature of criteria, by computing the probability of certain results, as if they were obtained merely by chance. The goal is to separate art from chance. At this point, with Alejandra, we find it necessary to discuss the binomial model.

5.2.2 How good is eight out of ten?

Once data are collected, the natural follow-up is to find a way to assess how strong is that result as evidence for the ability of that student to distinguish between juices. We focus on Diego's results, eight out of ten guesses, as an example to build on. As teacher and researcher, our goal is for students to compute and interpret a development as in Equation 5.2.

Under the assumption of independence between trials and equiprobability of success and failure ($p = 1 - p = 1/2$) on each trial, the number of successful guesses on $n = 10$ trials can be modelled as a random variable $X \sim Binomial(10, 1/2)$. If so, then

$$\begin{aligned}\mathbb{P}(X = j) &= \binom{n}{j} p^j (1-p)^{n-j} && \forall j \in \{0, \dots, n\}\\ &= \binom{10}{j} \left(\frac{1}{2}\right)^j \cdot \left(\frac{1}{2}\right)^{10-j} = \binom{10}{j} \left(\frac{1}{2}\right)^{10} && \forall j \in \{0, \dots, 10\}. \end{aligned} \quad (5.1)$$

I attempt to get in contact with students with a direct inquiry question:

RAIMUNDO: What is the probability of having obtained, out of 10 trials, eight or more correct? Moreover, to believe him, how should that probability be? [*awkward silence*] Very high, very low?

What follows is a sequence of rephrasing the question and shy monosyllabic unintelligible answers. I intend that, before doing any calculation, we should agree on the purpose of the calculation. After reviewing the video, I realise students avoid eye contact after the direct question of the type "what is the probability of...", to which I do not expect getting a straight answer. Instead, I reformulate to "how could we calculate the probability of...", without much initiative from students. Some type of scaffolding is needed.

We need to make a pedagogical decision with Alejandra. On the one hand, we want to pursue the inquiry approach focusing on students independent work and dialogical nature of classroom interaction (Alrø & Skovsmose, 2002). On the other hand, as Hmelo-Silver et al. (2007) clarify, inquiry learning is not the same as unguided discovery learning. This episode is a first example on how the structure given by the inquiry approach (Blomhøj, 2016) needs not to be taken as strict as to leave the clarification of mathematical content towards the end of the inquiry. Moreover, Alrø and Skovsmose (2002, p. 105) highlight the role of *identifying* as a dialogical feature which is "more clarifying than the process of locating, as it includes a *crystallising of mathematical ideas*" and "this provides a resource for further inquiry" (p. 13). We agree on taking the binomial model to the whiteboard. Alejandra takes the lead, and the relevance of her strategies can be illustrated by the recapitulation we make at the beginning of Session 3. She uses the analogy of tossing a fair coin repeatedly, as they are independent trials and all have the same probability of success and failure. The analogy is well understood, as this utterance at the beginning of Session 3 shows:

RAIMUNDO: We talked about tossing coins. What do coins and juice tasting have in common?
VALERIA: Tossing coins is heads or tails, and tasting juice is with or without sugar. Either correct or incorrect, the answer.

Using Pascal's triangle as a visual aid, she poses the exercise of computing binomial probabilities for all cases, with 2 and 3 trials, which students solve taking turns on the whiteboard. In order to extend to 10 trials, Alejandra begins posing the exercise $\mathbb{P}(X = 10)$, which students find no big trouble in solving, as it means multiplying the probability of success $1/2$ ten times. This procedure is confirmed during the recapitulation at the beginning of Session 3:

RAIMUNDO: So there are two possibilities, of equal probability, half and half, 50 and 50, $1/2$ and $1/2$. Remember we reached a much smaller result somewhere, of $1/1,024$. Where did that come from?
NAHUEL: From the...
VALERIA: Two to the power of – no, $1/2$ to the power of ten.
RAIMUNDO: Why the power of 10, now that we are at it?
NAHUEL: Because we did ten trials.
RAIMUNDO: And why is it the *power* of ten and not multiplied by ten?
VALERIA: Because $1/2$ is repeated ten times. I mean, $1/2$ is multiplied ten times.

However, when Alejandra poses the exercise of computing the binomial probability $\mathbb{P}(X = 8)$ (required for our P-value), students realise that there are many ways of obtaining eight successes. Valeria volunteers to construct Pascal's triangle for ten trials on the whiteboard, only to find herself overwhelmed by the exponential growth of branches, needed to count the number of combinations.

The other principle at stake for the rest of the session is *pragmatism*. On the one hand, the binomial model is on the 11th grade *curricular framework* (MINEDUC, 2015b). On the other hand, students do not master combinatorial computations as they are supposed to, so part of their *previous knowledge* was a strong assumption. With Alejandra, we agree to let students deal with combinations using Pascal's triangle, tedious though it is.

In hindsight, this decision was wrong. For once, it was excessively time consuming and prone to error, as we verified by supporting students while doing their calculations. Therefore, if *time availability* was a preoccupation, letting Alejandra do her art and settle the binomial model formally could have indeed saved time. Besides, the abstraction that a coefficient $\binom{n}{k}$ may give was curtailed, i.e. the problem became too local and difficult to generalise.

Here arises another tension between perspectives on exemplarity. On the one hand, for the sake of analysing the judicial case on Session 3, computing the P-value and interpreting its meaning regarding conveying evidence is enough to make it *critically exemplary*, when connecting it to the case of Malcolm Collins. However, if a stronger focus is put into *instrumental exemplarity*, then the abstraction to the concept of random variables and their distribution is undoubtedly relevant.

Despite the struggle, by the end of Session 2, each of the groups have computed the probability we aimed for. The corresponding P-value associated to Diego's result, i.e. the probability of obtaining a result at least as successful as that is the probability of a realisation of a binomial random variable being eight or larger. That is:

$$\begin{aligned}\mathbb{P}(X \geq 8) &= \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X = 10) \\ &= \binom{10}{8}\left(\frac{1}{2}\right)^{8} \cdot \left(\frac{1}{2}\right)^{2} + \binom{10}{9}\left(\frac{1}{2}\right)^{9} \cdot \left(\frac{1}{2}\right)^{1} + \binom{10}{10}\left(\frac{1}{2}\right)^{10} \cdot \left(\frac{1}{2}\right)^{0} \\ &= \left(\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right)\left(\frac{1}{2}\right)^{10} \\ &= (45 + 10 + 1)\left(\frac{1}{2}\right)^{10} = \frac{56}{1,024} \approx 0.0547 \end{aligned} \tag{5.2}$$

The meaning of the result on Equation 5.2 is more important, which is the last crystallisation I make during Session 2:

RAIMUNDO: What was that probability, in the end? We computed the probability of obtaining eight or more successful guesses. What result did we obtain?

VALERIA: $56/1,024$?

RAIMUNDO: And how much is that as a percentage, more or less?

CARLOS: 5%.

RAIMUNDO: So we are saying that the probability of obtaining a result as good as eight or more is around 5%, by doing it merely by chance. Does that give him credibility?

VALERIA AND PABLO: Yes.

RAIMUNDO: So it seems very unlikely to reach such a result by guessing using no ability. That is what we will try to see whether it is valid or not [on next session].

The goal of Session 2 was to draw on the data collected during Session 1 and interpret it as evidence, utilising probabilities. In this case, the idea was to interpret eight out of ten successful trials of a right/wrong experiment of tasting juice (data), by computing the probability of obtaining a result at least as good as eight by chance alone (evidence).

I take responsibility for some tensions between design principles we could have prevented. The inquiry approach should not be taken to the extreme of unguided discovery learning. Crystallising mathematical ideas, such as the binomial model in its general formulation, is necessary to follow the inquiry, whereas the alternative of counting combinations on a tree diagram is unnecessarily time-consuming, compromising the pragmatic restriction of time. It took the whole 90-minute session to reach the result on Equation 5.2 and no time was there for dialogue and reflections, let alone a generalisation of the binomial model as in Equation 5.1.

In fact, by re-watching the videos from Session 2, it seems evident that students lost the motivation they showed during the juice experiment on Session 1. In part, that experienced boredom may explain why four students did not show up on Session 3: the entire group C (Ayelén, Gabriela and Rayén) and Nahuel.

5.3 Session 3: From juice to justice

Session 3 starts by recalling what has been done during the previous session. In particular, we recall the calculations and assumptions behind the probability of achieving a result at least as successful as eight out of ten trials, namely $\mathbb{P}(X \geq 8) = 56/1,024$. Identifying its interpretation is more relevant:

RAIMUNDO: One phrase is to say: "we computed the probability of him having an ability, i.e. that he did not do it by chance". And the other is so say: "we computed the probability that, if he did it by chance, he obtains a result at least as good".
DIEGO: That is it.
RAIMUNDO: Which one did we do, the first or the second one?
VALERIA: The second one!
RAIMUNDO: Could there be a way of computing the probability of him having an ability? [*Silence follows*] It is strange, it is a strange question. How could I compute the probability of someone having an ability to distinguish? So the best we can do is to assume he does not. Let us assume he did it at random. What is the probability of him having at least such a good result? And that is what we computed yesterday.

After the recapitulation, we begin the discussion of the guiding inquiry on the title, namely the case of Malcolm Collins (People v. Collins, 1968). As mentioned at the beginning of this chapter, three elements are problematic in this case. Rosenthal (2015) refers to them as "when to multiply" question, "the prosecutor's fallacy", and the "out of how many" principle. The goal is to reflect upon those issues, in light of the judicial case and the notions of probabilistic evidence developed with the juice experiment.

The case of Malcolm Collins

On June 18, 1964, in a location of approximately 6,500,000 people,[7] an elderly lady was pushed down in an alley, and her purse was stolen. Witnesses said: a young caucasian woman, with a dark blond ponytail, ran away with the purse, into a yellow car, which was driven by a black man, who had a beard and moustache. Four days later, Malcolm and Janet Collins were arrested, primarily because they fit these same characteristics (at least mostly – Janet's hair was apparently light blond rather than dark blond).

At trial, the prosecutor called "a mathematics instructor at a nearby state college", whose identity no one seems to know. The prosecutor told the mathematics instructor to assume certain probabilities:

- Black man with a beard: 1 out of 10

- Man with a moustache: 1 out of 4
- White woman with blond hair: 1 out of 3
- Woman with a ponytail: 1 out of 10
- Interracial couple in a car: 1 out of 1,000
- Yellow car: 1 out of 10

The mathematics instructor then computed the probability that a random couple would satisfy all of these criteria, by multiplying:

$$\frac{1}{10} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{10} \cdot \frac{1}{1,000} \cdot \frac{1}{10} \cdot = \frac{1}{12,000,000}$$

It was thus asserted that there was just one chance in 12 million that a couple would have these same characteristics if they were not guilty. Malcolm Collins was convicted at trial, primarily based on this "one in 12 million" probability.

Adapted from Rosenthal (2015, p. 15), based on People v. Collins (1968)

After reading the text out loud, students check quickly that the multiplication there is correct. What comes next is a series of reflections in plenum about the case and connected to the whole activity.

5.3.1 When to multiply? The theoretical character of independence

From a mathematical theory of probability point of view, independence between two events A and B is the property by which the probability of their intersection is $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. This multiplicative equivalence is not just a property; it is its definition. If the assumption of independence cannot be assumed, by definition, those events are conditioned, and a solution to the "when to multiply" question can be to use a conditional probability, e.g. $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$. In his discussion about conditional probabilities, one of the issues raised by Borovcnik (2017, p. 1504) is the fact that the direction of conditionality between events in $\mathbb{P}(A|B)$ can be interpreted in many ways; "by time (earlier, later) by cause (cause, effect), or by indication (symptom, status)". These interpretations can be used to assess independence, but they are limited. After all, the meaning of independence is not trivial in real-world situations.

The binomial distribution model assumes the repetition of independent Bernoulli trials to be constructed. Our pragmatic design principle encouraged us to use the binomial model and, from that perspective, to find an example which could be assessed through it. That is how we came about the juice experiment. Certain intuitions are brought up by students during the experimental design in the quest for trustworthiness when tasting juice, as it was evidenced in Session 1:

RAIMUNDO: Each time he is correct, will you tell him?
VALERIA: Of course not!
RAIMUNDO: Why?
NAHUEL: Because I would familiarise with the flavour.

RAIMUNDO: So trials should be...? What word can we use?

The experimental design required repeated trials that did not have *affect* one another. If we tell Nahuel that he guessed correctly on the first trial, his guess on the second trial will be conditioned by the first one, in that sense. In fact, further improvement of the experiment was proposed by a group during Session 2, after Alejandra pointed out the independence assumption for the binomial model. One idea that crystallised was that of events not *affecting* each other:

RAIMUNDO: (...) The reason we could multiply had to do with the experiment you did. Especially your mates back there (GROUP B), insisted on a very important idea, which is that, if I try some juice and keep the taste [in my mouth], when I try the next one I can distinguish. Therefore, the result of one trial affects the next one. But, for being able to multiply probabilities, how should each of the trials be?

CATALINA: Independent.

RAIMUNDO: They have to be independent. Yes, correct. So they had an idea to correct that (GROUP B). What was that idea?

GROUP B AS A GROUP: Drink water [in between].

Once again, drinking water in between trials would prevent one tasting from *affect* the next one, since the student cannot contrast the current flavour with the previous one.

However, the assumption of independence is less intuitive when discussing the case of Malcolm Collins. Let us take, for example, the first two events given are "black man with a beard" and "man with moustache". A student notices that, for the prosecution's sake, being able to multiply them is convenient:

VALERIA: It says that he has a beard and moustache, and below those are taken separately.

RAIMUNDO: Sure, they (probabilities) are being multiplied, but when were we allowed multiplying? When events are...

ALL: Independent.

RAIMUNDO: Do you think having a moustache is independent of having a beard? (...) Could those characteristics be independent of each other? Or is it, for example, more likely that someone has a moustache knowing he has a beard?

VALERIA: They are independent.

MARTA: Those are different things.

As leading as my question is, the students find no conflicts with the assumption of independence since they are different variables. Here is a key disconnection between the juice experiment and the judicial case. In the first, independence between one trial and other appears as an alternative to a sequential cause-effect type of association (tasting once may influence the next), whereas, in the latter, independence can be scrutinised against any type of association. Hence my failed attempt to suggest that having a beard and having a moustache may be associated. Since having a beard is unarguably not a cause for having a moustache and vice-versa, students do not doubt the assumption of independence. This association is both a weakness and an opportunity in the sequence. What could have been done is precisely to provoke the debate on the distinction between causal and random perceptions of conditional probabilities, as Borovcnik suggests (e.g. Borovcnik, 2012, 2017). However, for the sake of time restrictions, I steer the conversation towards other events that are more evidently associated:

RAIMUNDO: If we already have the characteristics of "black man" and "white woman" identified, is "an interracial couple" independent from those two?

VALERIA: It is the same.

RAIMUNDO: We are already saying it.

MARTA: They come together.

In the local framework for analysis, these reflections refer to the "when to multiply" question (Rosenthal, 2015), but more comprehensive frameworks for probabilistic thinking would situate them as the notion of independence (e.g. Jones et al., 1997; Gal, 2005; Borovcnik, 2017). As for reflective knowing, students are not just engaging in mathematical reflections; neither are they reflecting upon how mathematical approaches shape our conception of some part of the world. Their reflections are technological, close to the third entry point on reflective knowing (Skovsmose, 1992, 1994): "Are the results reliable for the purpose we have in mind?" Students are reflecting upon the appropriateness of multiplication of probabilities for the purpose of conveying evidence, thus engaging in a technological reflection.

5.3.2 The prosecutor's fallacy: Probabilistic evidence, asymmetry of conditional probabilities

After data are produced, probabilities play the role of conveying evidence in the form of telling "art from chance" (Aburthnot, 1710; Hacking, 1975), i.e. as P-values. Even during the experimental design, the issue of reliability built an intuition from students to test for results more convincing than one which could be conveyed by a "fifty-fifty" chance, leading to repeated and justifiably independent trials. As the design of the experiment, collection of data, and calculation of the binomial probability $\mathbb{P}(X \geq 8) = \frac{56}{1{,}024}$ (Equation 5.2) were anchored in the question of credibility, the purpose and interpretation of the result are – to my surprise, I confess – non-problematic:

RAIMUNDO: What calculations did we do yesterday?

CARLOS: To see whether it was due to ability or by chance.

RAIMUNDO: We mentioned coin toss experiments (...) How is that related to our famous juice experiment?

CATALINA: What is the probability that the person's – in this case, DIEGO, who got the most right guesses – result or higher could happen by chance.

RAIMUNDO: Correct.

However, even if the P-value is well interpreted, Borovcnik (2017, p. 1502) argues that such "inferential methods are intended to preserve an FQT conception of probability so that these entities are interpreted in a frequentist sense" and it is one of the big problems highlighted in White and Gorard (2017) against the use of inferential statistics. The experiment that is repeated in the interpretation of a P-value is the entire study, and we do not make the thought experiment of repeating the juice study a large number of times. In the case of a judicial case, maybe a court can position itself applying a similar method to a large number of suspects and accepting a small Type-I error (ruled guilty when innocent), but from the innocent's point of view, this is impossible and the damage immeasurable.

Different meanings of probability can be conflicted. From a frequentist (FQT) perspective, a probability is interpreted as the relative frequency of a repeated experiment. From an *a priori* (APT) perspective, probability refers to a ratio between the multiplicities of favourable and total cases (Borovcnik & Kapadia, 2014). From a logical perspective (Carnap, 1950), probability is an objective degree of belief, a continuum between impossibility (0) and certainty (1), supported by evidence (Batanero & Díaz, 2007a; Batanero et al., 2016). Different meanings of probability enter in conflict when wrongfully interpreting the P-value as $\mathbb{P}(H_0|D)$ in general (White & Gorard, 2017) and as the "prosecutor's fallacy" in particular (Gigerenzer, 2002; Rosenthal, 2015). A way to understand this quite common fallacy is from two conflicting meanings of probability: the frequentist (FQT), as mentioned above, and logical. The juice experiment was constructed under the question of belief in the taster's ability to distinguish between sugary and non-sugary juice. Students find a similar driving force in the case of Malcolm Collins:

RAIMUNDO: What does the prosecutor want to prove, innocence or guilt?

ALL: Guilt.

RAIMUNDO: What is the prosecutor trying to do with that probability (1 out of 12,000,000)?

VALERIA: [He is] trying to compute how many people could be guilty.

MARTA: As an argument, if it is higher, then the probability of them being guilty is higher.

Marta is trying to identify by justifying that a high number (12,000,000), equivalent to a low probability (1 out of 12,000,000), is implying a high probability for the suspect to be guilty. But this probability $\mathbb{P}(\text{innocence}|\text{evidence})$ is unknown and simply cannot be deduced from $\mathbb{P}(\text{evidence}|\text{innocence})$ alone. In theory, it could be achievable by means of Bayes' formula:

$$\mathbb{P}(\text{innocence}|\text{evidence}) = \frac{\mathbb{P}(\text{evidence}|\text{innocence}) \cdot \mathbb{P}(\text{innocence})}{\mathbb{P}(\text{evidence})}, \tag{5.3}$$

but Equation 5.3 requires us to know the unconditional probabilities $\mathbb{P}(\text{innocence})$ and $\mathbb{P}(\text{evidence})$, for which there is no information (Gigerenzer, 2002, 2003). This is the essence of the prosecutor's fallacy, i.e. the perception to find a low likelihood for a suspect to be innocent, given the evidence gathered, whereas, in reality, he or she has found the likelihood of gathering such evidence under the presumption of innocence. It is, in a broader sense, the issue of asymmetry of conditional probabilities (Borovcnik, 2017).

Despite the correct interpretation of the P-value for the juice experiment, this fallacy or confusion is quite pervasive. Even towards the end of the joint reflection, when discussing the consequence of multiplying probabilities of questionably independent events:

RAIMUNDO: If each time we add characteristics which are redundant or non-independent...

VALERIA: It increases the number, but decreases the... eh...

ALL: Probability.

MARTA: The probability of innocence.

RAIMUNDO: The probability of innocence? What is the probability being computed there?

CARLOS: [The probability] of finding a person with those characteristics.

RAIMUNDO: Correct... if it was found by chance.

5.3.3 Out of how many? The theoretical character of probability

The probability of an event is often attached to a chance experiment, and its meaning is unclear when the experiment is unspecified or unclear. This aspect is the first highlighted in the framework of probabilistic thinking of Jones et al. (1997). Engaging students in the design of their own experiment seems a good start to realise that numbers do not speak for themselves, as data cannot be detached to context. Pablo's suggestion that "the entire group should try" is well aligned to the initial contextualisation given to the problem of whether products with and without sugar taste differently. Data would then be collected in the form of a survey and more appropriately be analysed accordingly. However, due to our pragmatic goal of covering the binomial distribution and a need to connect to the exemplary judicial case, we reformulate the question to a single person's ability to distinguish between juices. Here lies a potential for reflecting according to the third entry point (technological) on reflective knowing (Skovsmose, 1992, 1994): "Are the results reliable for the purpose we have in mind?" The purpose is ultimately to evaluate the difference in taste between sugary and non-sugary juice, but two approaches are conflicted: surveying a sample of different people once each, and having one person to taste repeatedly, as we did.

Discussing the case of Malcolm Collins, the first problem found interpreting probabilities in context is connected to the "out of how many" principle as defined by (Rosenthal, 2015). As soon as we finish reading the case out loud, Valeria reflects:

VALERIA: There is a number which is useless.

RAIMUNDO: What number is useless?

VALERIA: I do not know, but that $1/12,000,000$ is too big for $6,500,000$ people.

RAIMUNDO: OK. Describe... elaborate a bit more. [*To all*] She is saying that something seems strange. The probability is low, but there are many people in that city. What is the problem there?

VALERIA: In that city there are only $6,500,000$ people. But the probability is above the people in the city.

...

RAIMUNDO: What is the prosecutor trying to do with that probability?

VALERIA: [He is] trying to compute how many people could be guilty.

MARTA: As an argument, if it is higher, then the probability of them being guilty is higher.

Valeria is thinking aloud, manifesting her discomfort while interpreting the probability given as 1 out of 12,000,000, in a location of 6,500,000 people. While we locate (not identify) the issue of the "out of how many" principle, I try to follow the issue in the context of the purpose in mind. Marta is saying that a high number (12,000,000), equivalent to a low probability (1 out of 12,000,000), is implying a high probability for the suspect to be guilty. The discussion does not follow, but, potentially, the third entry point on reflective knowing could take place here. In context, the overall purpose can be interpreted as a logical meaning of probability (Batanero et al., 2016). With that purpose in mind, just as in the juice experiment, at least two different approaches can be taken: to find the probability of a random couple to match the criteria or the probability that *two or more couples* match those same characteristics. Out of, say, one million couples, this probability is much higher.

5.3.4 What the local framework does not say

The domain-specific framework used for analysis is used to identify issues of probabilistic thinking that are at stake for the judicial case described by Rosenthal (2015). Despite how useful it is to identify relevant probabilistic themes, these can be approached from broader perspectives on probabilistic thinking (e.g. Jones et al., 1997; Gal, 2005; Borovcnik, 2017), as they relate to conflicting meanings of probability, the theoretical character of independence and the correct probability of an event.

Additionally, these themes are viewed from a right-or-wrong perspective. The "when to multiply" question refers to an assessment of whether events are independent or not. The prosecutor's fallacy has its name for a reason. The "out of how many" principle can be seen out of the right-or-wrong realm, ye, so far it reveals actual and potential technological reflections. Are there critical reflections that could take place? Can students reflect upon the formatting power of mathematics and probability?

Let us take at a discussion about possible scenarios that may take place as a result of a trial:

VALERIA: They could say that the person is innocent when in reality she is guilty. And that the person is guilty, but in reality, she is innocent. Also that they say she is guilty when in reality is guilty. And that she is innocent when in reality is innocent.

RAIMUNDO: The latter two you said, are those risks?

VALERIA: No. Those are in case we reach the truth.

RAIMUNDO: That is what we would like to.

RAIMUNDO: Out of the two risks she mentioned, which one is worse?

VALERIA: To say someone is guilty when in reality he is not.

CARLOS: When he is declared innocent but was, in fact, guilty, that is also a problem.

RAIMUNDO: It is also a problem.

VALERIA: It is a risk because he will keep doing it.

RAIMUNDO: Both are risks. Do you think there is one worse than the other?

VALERIA: The one of saying he is innocent, but in reality, he is guilty. Because he will do what he did again.

DIEGO: I think the whole opposite.

PABLO: Yeah, me too.

DIEGO: I think if he is declared guilty, and he did not do it. It is worse.

RAIMUNDO: The truth is this is subject to opinion.

Valeria and Carlos first agree on the false-negative (type-II error) to be worse. Then Valeria changes her mind and, in principle, does not agree with Diego on which of the risks (type-I or type-II errors) is worse. The issue at hand is the notion of risk. Judgements of risk are rooted in context and depend not only on probability but on judgements of impact, which are highly influenced by context (Pratt et al., 2011; Levinson et al., 2012). Risks also are subject to values; they cannot merely be statements of fact (Beck, 2000). Which would be worse, to convict an innocent man or to let a criminal free? It cannot be calculated.

Moreover, each of the wrongful decisions in the unwanted scenarios is perceived much differently from stakeholders. For example, a type-I error could be seen from an FQT perspective for the court, as a low rate of wrongful convictions. From an innocent person, it can change his or her life irreparably. This asymmetry is one of the issues of the theoretical character of small probabilities and makes sense in scenarios of decision making under risk (Batanero & Borovcnik, 2016; Borovcnik, 2017).

Perhaps the closest reflection upon the formatting power comes after a direct prompting on my behalf, in the form of a hypothetical question:

RAIMUNDO: If all calculations are correct, do you think it is just to convict him based on this probability?
VALERIA: Sure.
PABLO: No.
RAIMUNDO: This is a matter of opinion. So [*to* PABLO] why not?
PABLO: Because there is still a possibility that he is innocent. There are still more people with those characteristics.

Once again, the issue of risk and small probabilities can be an input for critical reflections. There is no correct answer to my question. Valeria advocates for the (proper) use of probability as a just way to convey evidence for a conviction. Pablo challenges her, locating – not identifying – the problem of small probabilities (Alrø & Skovsmose, 2002). For him, the fact that the suspect still can be innocent, even if the P-value is well computed, interpreted, and very small, makes a conviction unacceptable. I cannot make a claim based on the evidence, but I dare to speculate that these interactions have the potential for students to acknowledge that something is problematic about using probability calculations to convey criminal evidence. Probability is doing something. It takes part in the formatting of our conception of what is just.

The case of Malcolm Collins reached the Supreme Court of California, where his conviction was reversed on the grounds of the non-transparent sources of data and grotesque miscalculations. The resolution makes a warning:

> Mathematics, a veritable sorcerer in our computerised society, while assisting the trier of fact in the search for truth, must not cast a spell over him. We conclude that on the record before us defendant should not have had his guilt determined by the odds and that he is entitled to a new trial. We reverse the judgement.
>
> (People v. Collins, 1968)

5.4 First considerations on the design principles

I have pointed out some challenges and tensions during the classroom experience that deserve to be revised in terms of the design principles, to begin addressing **RQ3**.

The *inquiry approach* should not be mistaken by self-discovery learning, especially when the inquiry is open-ended (Hmelo-Silver et al., 2007). In Session 2, making a pause for exercising probability calculations was necessary to move on with the inquiry. Therefore, the joint reflection and establishment of mathematical knowledge (Blomhøj, 2016) needs not to be the very last part of the inquiry work, but a constant adaptive way of supporting students' learning.

The use of two inquiries in one learning environment was meant to make connections from one to another. The juice experiment was time-consuming, to the lengths of taking two 90 minute sessions from start to finish, from experimental design to computation and discussion of the P-value. It left us 45 minutes to discuss the judicial case, where, besides, four fewer students participated. Given the *pragmatic* time restrictions, an improvement of the task could rely on one single *guiding inquiry*. The case of Malcolm Collins is *authentic* as an event, and so are the data, if one takes it at face-value as presented by the prosecutor. If one would turn the case into an inquiry for students to develop, the availability of data can be a challenge, so the use of fictitious information may be an alternative, but it could compromise the authenticity of the task.

Thus far, exemplarity has played the role of assessing the quality of the example, from different perspectives. However, exemplarity should be lived up. The probability computed as Equation 5.2 is an example but is not exemplary unless the more general notion is established in the plenum. The same can be said about the critical case of Malcolm Collins. If it is meant to be exemplary for how mathematics shapes part of our world, a more general stance towards probability and justice could have been made. Some interesting discussions take place while analysing the judicial case, where students reflect upon miscalculations and misinterpretations, possible risks (type-I and type-II errors), and whether Malcolm Collins should have been convicted. However, the case turns exemplary only once they discuss in a broader sense, for example, whether it is worse to convict an innocent person or to let a criminal free. This connection is related to the debate between situated learning and transfer. Advocating for situated learning presumes that cognition is context-dependent and skill transfer or abstraction often fails to be achieved. However, successful stories of transfer and abstraction are documented as well in empirical studies (J. R. Anderson, Reder, & Simon, 1996). The point is that the question of transfer is not binary. Instead, the goal is to find ways to teach for transfer (Niss & Blum, 2020). Consequently, exemplarity should be seen no only as a quality of the cases, but also as a quality of the teaching practice.

The learning environment of this chapter makes use of two examples, namely the juice experiment and the case of Malcolm Collins. These are *exemplary* from different perspectives. The former is intended to be instrumental for disciplinary ideas, such as the binomial distribution and informal hypothesis testing (Zieffler et al., 2008; Lee et al., 2010). Students design and experiment themselves, living up to the subjective perspective on exemplarity, since they can draw from their own experiences. The latter is exemplary in the critical sense, as an application of probabilistic ideas to shape notions of evidence and justice. These versions of exemplarity enter in tension when making pedagogical decisions. Crystallising the binomial distribution in a general sense for the juice experiment is necessary for making an instrumental example. However, for the sake of the judicial case, it is not. The multiplication of probabilities of independent events is sufficient to address the “when to multiply” question (Rosenthal, 2015).

Furthermore, analysing the case of Malcolm Collins does not involve students personal experiences. By analysing a case, students are not involved in the inquiry; they are spectators or consumers of probabilistic and statistical messages (Gal, 2002, 2005). Students are not given the opportunity to make their investigation and decision on the judicial case. Enabling students to be producers of messages is a central idea to move forward in the following chapters.

Overall, the challenges illustrated in this chapter can be approached by a single authentic inquiry, making curricular adaptations according to students' previous knowledge. It may help to choose an example that is more connected to students' personal experiences and to give opportunities to produce their own statistical messages and decisions, beyond analysing and evaluating those produced by others. This step is what we attempt in the next chapter.

Notes

[7]The case took place in Los Angeles County, CA. I decided to give away the number of people to avoid a tedious search about what a county is and its population at the time, and the confusion from the fact that Los Ángeles is a city in Chile, as well.

Chapter 6

"How are PSU scores computed?"

RAIMUNDO: Imagine that only you as a class were taking the PSU test. Or just one group...
DAVID: [*Joking, to* BRUNO] I wish you do bad [on the test] then, so I can do well.

PSU (*Prueba de Selección Universitaria* in Spanish) is the name given to the nation-wide standardised test that high school graduates take in Chile in order to apply for higher education study programmes. PSU scores are reported using a normalised relative scale, and each year's results lead to heated political debates using different sorts of statistical analyses and representations. It is a critical topic to explore with high school students, whose mathematics study programme includes the normal distribution in 12th grade, the same year they ought to take the national admission test PSU.

One of the expected learning goals on the 12th-grade curriculum is to "critically evaluate statistical information extracted from the media, such as newspapers and magazines, or the internet" (MINEDUC, 2015a, p. 87). In fact, the official study plan provides an exemplary activity of analysing a fictitious series of PSU score averages for a school throughout several years (MINEDUC, 2015a, pp. 88–89). Suggestions include the discussion of some of the issues in the graphical representation (e.g. lack of origin in the axes system) and the use of the coefficient of variation to analyse how significant the changes are. With Alejandra, the teacher, our purpose is to go deeper by contextualising the actual political use of such a diagram portrayed in Figure 6.3 and having students understand what PSU scores are actually indicating since the normal distribution is part of the curriculum as well. Thus, the general inquiry is given as the title of the chapter: "How are PSU scores computed?"

We count with three sessions of 90, 45 and 90 minutes, respectively. The sequence is structured accordingly:

1. Setting the scene by reading a news article of PSU results of the previous year, sharing the general inquiry and analysing an official conversion table (Table 6.2) to get an understanding of the percentile-based computation of PSU scores.

2. Computing PSU scores based on students' latest practice tests, according to three different cohorts: their working group, their whole class, and the previous year's national cohort.

3. Evaluating a graph of PSU score trends depicted on a medium (Figure 6.3) by focusing on technicalities (e.g. axes) and its interpretation based on an understanding of PSU scores.

Before the implementation of the task, the overall purpose was to provoke critical reflections about politically framed statistical information, as I report in Section 6.3 concerning the third part of the sequence. The first two sessions are meant to set the scene and enrich the quality of the critical evaluation of Figure 6.3, which they do. In hindsight, some unexpected topics of the formatting power of statistics emerge during Sessions 1 and 2, as I report in Sections 6.1 and 6.2. These issues are not unfolded with the students but from them. They do represent the potential of the inquiry for developing further reflections if replicated in the future.

6.1 Session 1: Percentiles in the class-room, socioeconomic background and educational foreground

PSU consists of a battery of four multiple-choice tests; two of them mandatory to participate in the admission process – Spanish Language and Mathematics – and two of them optional – Natural Science, and History and Social Science –, depending on the programmes and institutions each student intends to apply for. Each of the test results is given on a scale from 150 to 850 points. Students apply with an admission score computed as a weighted average of all relevant tests and their high school grades. Each institution-programme decides these relative weights.

PSU is a hot topic for 12th-grade students. That is why this inquiry was included among my suggestions to Alejandra. She is the high school mathematics teacher for a relatively small school in a rural city of 17,000 people, where students in 11th-12th grades can follow either a vocational-tourism, vocational-nursing or humanistic-scientific programme. Mathematics is a core subject in all programmes. However, vocational programmes count with only three pedagogical hours (135 minutes) per week, as opposed to the humanistic-scientific programme that counts with five pedagogical hours (225 minutes) of mathematics per week plus complementary mathematics courses. With Alejandra, our choice is to work with her humanistic-scientific class, not only on the grounds of time available but also of the focus on higher education. That is, we chose a class for which PSU is supposed to be more appealing as a topic of discussion.

It is a class of eight students, so most discussions take place in the plenum. Nonetheless, they are distributed in the classroom into two groups. One group includes Néstor, Claudia, Bruno, David and Miguel, who does not show up during Session 1. The other group includes Andrea, Laura and Matías.

6.1.1 Background

After introducing myself and the overall plan for the activity, students begin by reading out loud a news article about PSU results form the previous year (Tele 13, 2017):

> *December 26th, 2017*
> **PSU 2017: 3 out of 10 students do not reach the minimum score to apply**
>
> Out of the 261 thousand students taking the test, 79,948 obtained less than 450 points. Only 213 obtained more than 800 points.

> "Here is a socioeconomic gap that is brutal" stated this morning the minister of Education, Adriana Delpiano, after analysing the PSU 2017 results (...)

The article goes on displaying numerical facts about the distribution of students in different ranges of PSU scores according to the type of school (public, subsidised or private), appending Table 6.1.

The leading learning goal is to evaluate statistical information from the media critically. One of the bottlenecks agreed by researchers regarding statistical literacy is number sense, in particular, the use of rational numbers in different forms (e.g. fractions and percents) (see, e.g. Joram, Resnick, & Gabriele, 1995; Gal, 2002). By taking Skovsmose's types of knowing (Skovsmose, 1994, Chapter 6) as a way to guide the conversation, I decided to begin by triggering mathematical knowing. Three fundamental skills are necessary to be able to understand the official conversion table (Table 6.2) and compute PSU scores: computing percentages, extracting information from a table and working with percentiles.

First, we ask students to verify the information on the heading, i.e. to simply verify that the following proportion holds: Number of students with less than 450 points : Total number of students $\approx 3 : 10$. The quantities in question are both given straightforward in the subheading of the news article and on Table 6.1, appended to it. It goes as follows:

RAIMUNDO: That is what I want you to do in your groups. The heading says "3 out of 10 students do not reach the minimum score to apply". Where do those number come from? How do you compute that number?

NESTOR: From the total. And then you start separating in groups of 10...

DAVID: From the total there should have been a percentage. And that must have been three point something, and then...

...

ALEJANDRA: Come on, apply what you already know!

BRUNO: Is this statistics?

ALEJANDRA: Yes.

...

DAVID: Is this what we did on the computers with Excel?

ALEJANDRA: [*Smiles at them with calm.*]

BRUNO: We should multiply by a number, then sum all of them together and then divide by the grand total.

ALEJANDRA: [*Raises her eyebrow sceptically*] That is the average.

BRUNO: But that is the average, indeed!

Admittedly, I expected a quick calculation to come up. For example, students could have divided $79,948/261,987 \approx 0.3051$ with their calculators. From there, they could make an association – whether intuitive or computed – from the decimal number 0.3 to the ratio "3 out of 10" or the percentage "30 per cent". However, several confusions arise in the early discussion as students think aloud.

Students are mostly locating – not identifying – (Alrø & Skovsmose, 2002) possibilities to solve the task. That is, they try things out without much justification. Néstor begins thinking about dividing the cohort into groups of 10, as in "3 out of 10" literally. David thinks in percentages, but his intuition is inverse to what could be a useful thought, i.e. that in the total

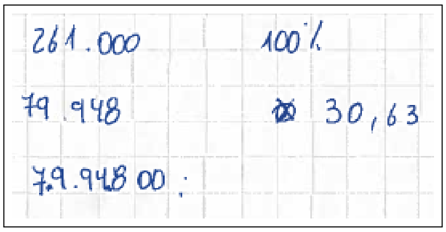


Figure 6.1: Andrea's verification of 30%

cohort, the percentage of students in that category is "30 point something" and thus equivalent to "3 point something" out of 10. Bruno realises that similar exercises have been done in statistics and David reinforces the idea remembering calculations done in Excel. This aspect of the didactical contract (Brusseau, 2002) puts him in a disposition to do what he is expected to do within statistics and suggests to do the same as they remember doing before; apparently, computing the "average" (sample mean, actually) of a variable form a frequency table, which is not the exercise we have in mind.

Later on, students remember that percentages are solved by arranging them as a "rule of three", thus identifying the proportional nature of percentages. The solution by Andrea in Figure 6.1 displays this arrangement, where the unknown x is then replaced by the solution 30.63%. The group initiating the discussion above had more difficulties:

BRUNO: (…) It gives [*types in calculator*] 325. That is too much. I did something wrong.
ALEJANDRA: Type it again.
BRUNO: It would be 100... 79,948 divided by 100. No, it does not work. Then it would be 100 divided by 79 thousand... No.
ALEJANDRA: That neither.
NESTOR: Times! It is *times* 100 (…) Oh, I know how to do it!
ALEJANDRA: Is it then multiplied again?
BRUNO: Divided [*types in calculator*]. There it is. [*Eureka face*] Aaaah!

Neither Alejandra nor I give away the procedure, but support them to reconsider and correct their calculations. Mathematical knowing, evident in mathematical discussions (Barbosa, 2006) can be observed among students' questions of the type "are we doing the calculations right?" and "are we doing the right calculations?"' (Skovsmose, 1994, Chapter 6). The first question type is evidenced by Bruno realising something must be wrong with the result and Alejandra suggesting to type it again in his calculator. The second type comprises the rest of the exchange by discussing what is to be multiplied and divided.

Approximately 15 minutes went by from the moment I posed the instruction to David's Eureka moment. It is a 12th-grade classroom, and I become worried about their educational background in mathematics. In our meeting after the session, Alejandra concurs. She is new as a teacher in this school and shares her surprise.

Nonetheless, the following step is to check their ability to interpret and extract information from tables, as it will be critical to follow through the rest of the sequence. Bruno and David display some of the challenges of interpreting information on the column labelled as "no info." from Table 6.1:

BRUNO: [It shows] how many people obtained a certain score.

Score range	No info.	Private	Subsidised	Public	Total
Less than 450	953	1,775	38,403	38,817	79,948
450 to 499.99	656	1,968	27,960	19,567	50,151
500 to 549.99	497	3,294	28,634	14,322	46,747
550 to 599.99	342	5,112	23,391	9,630	38,475
600 to 649.99	227	6,248	14,086	5,391	25,952
650 to 699.99	75	5,002	5,772	2,366	13,215
700 to 749.99	20	2,839	1,873	892	5,624
750 to 799.99	7	1,045	403	207	1,662
800 or more		144	37	32	213
Total	2,777	27,427	140,559	91,224	261,987

Table 6.1: Summary of PSU 2017 results by type of school. Source: Tele 13

DAVID: But not all of them got the same score, for example...

BRUNO: For example, out of, let us say, 1,000 people, 953 obtained 450 points. So how many people got... [*unintelligible*]?

DAVID: ... between 450 and 4[99.9]. That is what I was saying!

BRUNO: It is the interval, then.

ALEJANDRA: But mind the whole table. You said "out of 1,000 people, 953 obtained [450 or less]". Where are those 1,000 people?

BRUNO: It is just to say something!

ALEJANDRA: Then look at the table. Look at the table.

Bruno identifies one variable categorised and counted, namely PSU score. David notices and explains to him that the PSU score variable is expressed in terms of intervals. The column total is not found yet, and Alejandra insists on the fact that the table has all the information needed. After the exchange above, a silence follows, David gets up and asks Alejandra to clarify this to him, which she does effectively.

Extracting single pieces of information from the table results more natural for them after discussing the overall outlook of Table 6.1:

RAIMUNDO: In which score-range the fewest students are found?

NESTOR: From 700... No, of 800 or more.

RAIMUNDO: And [within the same range], after "no information", where are the fewest students?

UNISON: Public!

NESTOR: There are 32.

It is time to move to Table 6.2, which is at the core of the guiding inquiry. It is the official Mathematics PSU conversion table used for the 2017 version, i.e. for the admission process to higher education of 2018. The table has three columns, corrected score, percentile and PSU score. The corrected score is the number of correct answers a student may get. The percentile, rounded to the closest integer, is the percentage of students in the cohort that obtained the corresponding corrected score or less. The PSU score, computed form the percentile of achievement, is the normalised reported score used in the admissions process. Now that students have proven to be able to work with percentages and tables, we need to develop the notion of percentile:

RAIMUNDO: What does percentile mean? Why [next to 850 points on Table 6.2] does it say 100?

LAURA: Because it is 100% of [correct] answers...

RAIMUNDO: (...) As your teacher told me when I spoke with her, you have been to talks on [student] grants and loans, right? There is a word, similar to percentile...

BRUNO: Quartile!

CLAUDIA: Quintile!

After asking a straightforward inquiry question to get in contact, Laura's answer makes me realise that the selected row is confusing and I shift to make connections with their personal experience and build up from there, defining a turning point (Alrø & Skovsmose, 2002). Although our interest is to develop the notion of percentile as a measure of position in an ordered cohort, the conversation takes an interesting shift towards socioeconomic categories:

CLAUDIA: For grants, quintiles are used.

...

LAURA: It is used to give grants to some and not to others.

RAIMUNDO: (...) That is fine, [but] it is the definition [what I am fishing for]. What does it mean that someone is within the first quintile?

CLAUDIA: To be of low [economic] resources. (...) Yes, but it is not related to the number of people you live with. It is [about] how are you doing economically, with the family.

...

RAIMUNDO: Why do grants apply for the first quintile? What is the last quintile? How many quintiles are there? [*awkward pause*] It is in the word[8]...

NESTOR: Five.

RAIMUNDO: (...) Grants go to the first quintile, why not the fifth quintile?

CLAUDIA: Because there lies society's highest rank. They have more money. They have the means to pay for their studies.

RAIMUNDO: So... they are arranged in order.

Claudia is well informed about conditions to apply for student grants and loans for higher education and how the division of families into income quintiles plays a significant role. Skovsmose (1998) argues that context-oriented reflections can be considered critical reflections, which Claudia is putting in evidence. I start a conversation to develop the notion of percentile amongst the students, at the service of the guiding inquiry. However, by contextualising the use of quantiles in their personal experience, model-oriented reflections are elicited by referring to the use of quintiles for grants and loans applications. By following up, Claudia even evokes the social context, i.e. the division and sorting of society into lower and "higher ranks". The latter reflection arises after an identifying dialogical feature, namely a why-question (Alrø & Skovsmose, 2002), which leads Claudia to justify why the higher ranks are not applicable for grants.

The mathematical-knowing background for responding to the central inquiry is so far established. Students have shown they can compute and interpret percentages, interpret and extract information from a table, and the ground is paved for working with percentiles.

Anchoring to context and personal experience allowed for reflective knowing to arise. It reveals an alternative understanding of backgrounds, beginning to confirm the subjective quality of the example. First, students find difficulties to compute a simple percentage, despite being a scientific-humanistic 12th-grade class. Second, extracting data from Table 6.1 could show how associated is PSU performance with the type of school (public, subsidised or private). Finally, quantiles are exemplary of the use of distributions to divide and rank society as kinds of people defined by their income. The session has the potential to make students reflect on the role of statistical information in making sense of their educational and socioeconomic background.

6.1.2 Foreground

As a result of reading Table 6.1, I decide to have a chat with students to set the scene of the topic. From my experience both as student and teacher, the questions "what do you want to study and where?", "what was last year's admission cutoff PSU score?" and "how are you doing on practice tests?" become small talk mantras among 12th graders. I had taken their aim for higher education for granted:

RAIMUNDO: Just to get a better understanding of the context, everyone here wants to study after high school, right?
[*Some students nod, others do not react, and* DAVID *shakes his head in negation.*]
RAIMUNDO: [*Rephrasing*] (...) What plans do you have for next year?

As it turns out, four of them intend to follow studies at higher education institutions whose admission is subject to PSU scores: Bruno is deciding between civil construction and obstetrics, Claudia wants to become a social worker, Matías is interested in pedagogy, and Néstor aims for studying for accountant-auditor. On the other hand, David is interested in joining the army, Andrea is volunteering for the military service, and Laura wants to be admitted to the police academy. That is, three of them only need to take the PSU test and their scores are not taken into account for admission (Abujatum, 2018). In principle, there is nothing problematic concerning students' personal choices about their future. Right?

Skovsmose (2005a) argues that learning obstacles are often referred to students' social and cultural backgrounds, and thus objectified and individualised. As a response, the notion of *foreground* has been suggested to conceptualise a more systemic approach on students' perspectives and interests. Foregrounds are "defined through very many parameters having to do with: economic conditions, social-economic processes of inclusion and exclusion, cultural values and traditions, public discourses [and] racism" (Ernest et al., 2016, p. 10), taking into account learning obstacles as subjective and political. However, foregrounds are situated and illustrated by the person's perception of possibilities: "the foreground is that set of possibilities which the social situation reveals to the individual" (Skovsmose, 1994, p. 179).

From a research perspective, the challenge becomes to deal with the dialectic between individual and structure (Wedege, 2016). Alrø, Skovsmose, and Valero (2009) propose two main principles for the use of interviews as inter-views, i.e. "seeing together". Firstly, "an active interviewing by the researcher need not be seen as a disturbance of what the foreground 'really' is" (p. 20). Second, the purpose is not to find *objective* notions of foreground, and thus a dialectical approach is recommended. Wedege (2016) builds on these ideas and connects them to sociologist Pierre Bourdieu's concept of *habitus*, in order to account for durable and transposable *dispositions* towards education. These dispositions are certainly conditioned by objective factors (background), "without being, by no means, the product of obeying rules they are collectively orchestrated without being the product of the organising action of a concert leader" (Bourdieu, 1980, p. 89, as cited in Wedege (2016)). The key is to interpret individual perceived possibilities as a product of a non-intentional systemic set of habitual dispositions in a dialectical fashion.

The PSU learning environment, in particular Session 1, has a potential to investigate the notion of foregrounds, since students have the opportunity to explore their educational performance in a test in the context of systemic inequalities. Moreover, the objective notion of background and subjective perception of foreground seem to clash within the dialectic between social structure and individual, as can be illustrated by some features of PSU as a topic, the statistics of order involved, and the students' close relation to the topic.

On the one hand, background elements are brought up on the table, starting by the simple verification of the information on a headline "PSU 2017: 3 out of 10 students do not reach the minimum score to apply" (Tele 13, 2017). It is, to Alejandra and me, surprising and discouraging that 12th-grade students face so many challenges just to compute a proportion correctly. Computing percentages and proportions are within the most basic type of questions in the Mathematics PSU test.

The same news article quotes the at-the-time minister of education, acknowledging the socioeconomic achievement gap in PSU (Tele 13, 2017). This claim is grounded from a public policy point of view, as Chile's school system is divided into public, subsidised and private schools. Unsurprisingly, this market-oriented school system makes the distribution of students among school type highly correlated to their socioeconomic background (Valenzuela, Bellei, & de los Ríos, 2014). The classroom a student partakes in turns into a class-room (Andrade-Molina, 2017).

During the scene-setting for computing PSU scores, I fish for an understanding of quantiles in general, and percentiles in particular, in order to interpret Table 6.2. Students are familiar term "quintile" as it is fundamental to know whether a high school graduate has the right to economic aid for higher education. They know this because talks about loans and grants are, logically, targeted to students from a low-income background. Furthermore, students acknowledge that label. This aspect of education *by the numbers* is problematised by Popkewitz and Lindblad (2018) as an aspect of statistical reasoning in educational policy:

> (...) the making of kinds of people inscribes a continuum of values and double gestures that normalises and differentiates in the efforts toward inclusion. While seeking inclusion, the very principles that are generated for inclusion divide and render certain groups as different, dangerous and in need of intervention."
>
> (Popkewitz & Lindblad, 2018, p. 217)

Popkewitz and Lindblad speak from the perspective of international educational assessment, where categories defined by ethnic background, migrant status, gender or income are defined with the good intention of making sense of educational achievement gaps, in order to direct policies towards filling them. At the same time, these labels may have an effect of perpetuating prejudices against certain "kinds" of people (Gutiérrez, 2008). The logic of identifying and targeting economic aid in terms of income quantiles can produce double gestures of (in)exclusions, one of the preoccupations of Critical Mathematics Education connected to issues of social justice.

On the other hand, students' expressions of educational foreground relate to individual aspects. After verifying the computation of the approximately 30% students who did not reach the minimum score, I ask students about their impression about such information. I intend to generate a sensation of *scandal* to counteract on the next session, but it goes in another direction:

RAIMUNDO: What do you think about that?
DAVID: That is their own responsibility [*General laugh*] What else could I say?
BRUNO: [*To* DAVID]But PSU does not measure your intelligence. PSU evaluates content from 9th grade. Try to remember content from 9th grade!
DAVID: See? It has to do with intelligence.
...
ANDREA: It is just that not all of us are intelligent. Not all of us will remember that content.
DAVID: One will always remember more what caught our attention.
...
NESTOR: Interest!

To explain low performances in PSU, students point to individual features, such as "personal responsibility", "intelligence", memory to "remember" content and personal "interest". For them, individuals shape the structure and not otherwise, as the notion of *habitus* could characterise (Bourdieu, 1980; Wedege, 2016).

However, the dialectic between individual and structure has political roots. According to Cavieres (2011), the focus on individual productivity is a consequence of the neoliberal educational reforms taking place in Chile since the 1990s. It culturally affects low-income groups sense of self-worth, i.e. "by stressing the notions of human capital and quality education, the reform has tended to reinforce these students' fatalism, and limit the scope of the organisations they form to improve their academic and social opportunities" (Cavieres, 2011, p. 1). These beliefs are pervasive, as Andreas Schleicher from OECD suggested. In an interview, he compares what students in Chile and Shanghai believe about their performance in mathematics: "while in one case they felt helpless with the educational system, in the other they were convinced about owning their success, so it should not be surprising that in the second case each student graduates with a solid mathematical base" (Revista de Educación, 2015). Whether manifest or hidden, structural factors mingle with personal expressions of students' foregrounds. Let us take, for example, David's justification for not following a higher education programme:

DAVID: I do like some careers, but why would I spend on something if I do not like studying? That is my point of view.
DAVID: I will start the [military] service and then see what I can find therein. If I like anything, I will specialise. And then my idea is to join the [army's] school for sub-officers or the Investigation Police.
DAVID: But this is to assure my future, more than anything. They (armed forces) get everything paid for. Even healthcare is for free.

At face value, students' foregrounds are explained by free-of-choice personal interests. David's comments, however, reveal some systematic issues related to his options, namely the "spending" needed in higher education programmes and the fact that the armed forces have free access to healthcare. Some of his classmates were also opting for joining the police or military, despite being part of a scientific-humanistic high school programme. It is possible that, had we discussed the relation between the type of school (public, subsidised or private) and income, they could have connected the dots between the information given on Table 6.1 and how their foregrounds had come to be. For example, the table can be expressed not in absolute by relative frequencies to illustrate the association between type of school and PSU score range, at least with an informal approach, since χ^2-tests for contingency tables would be far beyond the curricular scope.

Confronting the preoccupations of CME in the classroom is central to the general *problématique* is this thesis, and the notion of *foreground* is certainly appropriate for this learning environment. However, how would students interpret the systemic forces that produce their dispositions and perceived educational possibilities? Would that reinforce students' fatalism (Cavieres, 2011) and helplessness (Revista de Educación, 2015) after seeing themselves as a different and endangered kind of people (Popkewitz & Lindblad, 2018), who take part of a well-defined classroom (Andrade-Molina, 2017)? Is a statistical education that reveals the underlying political structures of learning obstacles worth it? It is an educational dilemma to which I do not possess a sharp answer. One possible exit to the dilemma is realising that the goals of education are far from simply performing in a standardised test, as we intend with this learning environment. The problem is, because of the same political structures, as David admits:

DAVID: In the end, PSU measures your future.

6.2 Session 2: Fabricated normality

> The normal stands indifferently for what is typical, the unenthusiastic objective average, but it also stands for what has been, good health, and for what shall be, our chosen destiny.
>
> (Hacking, 1990, p. 168)

Every time I have the chance to teach statistics and the normal distribution comes up as a topic of study; I like to ask students what is their guess for its name's coinage. Why is the Gaussian distribution called normal? Unanimously, answers are of the type "because many/-most things in nature follow this distribution". Students believe that the normal distribution resembles a natural order, normal in the sense of commonplace. During the rest of the course, they struggle to see that data sets they ought to work with do not fit a normal distribution, desired for applying parametric tests and related statistical methods. Transformations of data taken from a hat and arguable elimination of outliers become the order of the day. Collected data from their own experiments are no better, as their friendly emails asking for statistical aid reveal on following semesters.

It is not that data representing natural or social phenomena *do* tend to follow a normal distribution, but we *want* them to.

6.2.1 Social statistics

As it is characteristic of the shift towards modernity, methods used in the natural sciences began to be applied to make sense of and shape the social world (Davis & Hersh, 1986). In Section 3.1

I discussed the emergence of social statistics, wherein *statists* began to apply, for example, Gauss's distribution of errors in astronomical measurements to describe and assess a variety of biological, social and moral issues; a perception about human phenomena to be "as real as cosmic forces" (Hacking, 1990, p. 170). An iconic case is the notion of *l'homme moyen* (the average man) by Adolphe Quetelet, who measured the chests of five thousand Scottish soldiers, and discovered the bell-shaped curve when plotting his measurements, centred around their computed mean (Herschel, 1850). For Quetelet, "it is (...) almost as if Nature had aimed at producing this type – the ideal value for the group (...) [but] soldiers for different 'accidental' circumstances in their lives failed to realise fully the ideal standard" (Donnelly, 1998, p. 236). A duality lies in normality as a statement of natural fact, inclusive within the group around its mean, and the normal as an ideal, the Aristotelian mean, opposed to what is pathological. Normality is seen as a fact and as a goal.

Problematic uses of statistical models in the social world reside in the psychological roots of mathematics education research as well. Alfred Binet, cited in the premiere issue of *L'Enseignement Mathématique* (Binet, 1899), is known for his contributions to developing the first intelligence tests, though his studies discarded physical measurements to be explanatory (Kilpatrick, 1992). On a parallel track, eugenicists like Francis Galton attempted to apply Darwin's theory of evolution to interpret intelligence as a hereditary capacity (Desrosières, 1991). However, his use of statistical methods lacked rigour and are exemplary of a self-confirmatory logic (Hamilton & Rose, 1980). Despite Binet's insights, Kilpatrick (1992) recalls how some psychologists "developed out of Binet's tests a hereditarian theory of IQ that not only had some disastrous effects in its consequences for social policy but also coloured the views of a generation of American researchers in mathematics education about the prospects for improving mathematical abilities" (1992, p. 8). In later days, a best-seller book by Richard Herrnstein and Charles Murray makes (mis)use of statistics to make claims regarding intelligence throughout American society, including its hereditary ties to race. Its assumptions: "intelligence (...) must be depictable as a single number, capable of ranking people in linear order, genetically based, and effectively immutable" (Gould, 1994). Its take on an immutable, natural order of intelligence is given away on the title, *The Bell Curve*.

Modelling social, moral and psychological matters as normally distributed carry the historical baggage of the meanings given to normality. As a case of educational assessment, what would normality in PSU scores indicate? Could it the natural order of things? A normally distributed variable is undoubtedly friendly to work with as a statistician, and PSU scores would be no exception, for the sake of comparison and educational policymaking. However, if they do not follow a normal distribution, can that be fixed?

6.2.2 Tautological news

Let us come back to the inquiry. How are these scores computed after all? PSU test results are given in a scale from 150 to 850 points, according to the corrected score[9] (number of correct answers), but it is not through a linear function. For each subject (Spanish, Mathematics, Natural Science, History and Social Science), the PSU score is adjusted in such a way that they are normally distributed with a mean of 500 points and standard deviation of 110 points, fixing 150 points as the minimum and 850 points as the maximum (Antivilo, Contreras, & Hernández, 2015, p. 4). Therefore, PSU scores are computed as follows. Let us call CS the corrected score, let Z be an auxiliary random variable such that $Z \sim \mathcal{N}(500, 110)$, CS_{min} the minimum CS obtained by the cohort, and CS_{max} the maximum CS. Then

$$\text{PSU score}(CS) = \begin{cases} 150, & \text{if } CS = CS_{min} \\ 850, & \text{if } CS = CS_{max} \\ z \in \mathbb{R} : \text{Percentile}(CS) = \mathbb{P}\{z \leq Z\}, & \text{otherwise.} \end{cases} \tag{6.1}$$

Two aspects stand out from Equation 6.1. The first is that PSU scores are normally distributed by construction, by introducing $Z \sim \mathcal{N}(500, 110)$. The second is that the PSU score is a function of the corrected score only through the percentile and, as such, it depends on the corrected scores of the whole cohort. A student's PSU score does not tell her how well she performed. It tells her how well she performed compared to everyone else. PSU scores are given on a relative scale.

The headline of the news article shown to the students at the start states – in a scandalous manner – that approximately 30% of students did not reach 450 points, as PSU average between Spanish and Mathematics, which is the minimum required to participate in the admission process. The headline was shared throughout social media with shock by journalists and pundits. However, the proportion of students given less than 450 points is, by definition, the probability for an instance of a variable $Z \sim \mathcal{N}(500, 110)$ to result below 450, that is:

$$\begin{aligned} \mathbb{P}\{\text{PSU score} \leq 450\} &= \mathbb{P}\{Z \leq 450\} \\ &= \mathbb{P}\left\{\frac{Z-500}{110} \leq \frac{450-500}{110}\right\} \\ &= \Phi\left(\frac{450-500}{110}\right) = \Phi(-0.4545) \\ &= 0.3247 \approx 32\% \end{aligned} \tag{6.2}$$

Therefore, the headline merely says that 30% of students are among the 32% least-achieving students, what can only be characterised as tautological news.

I do not claim that this fact is a conspiracy. PSU is a nation-wide admission test, and "selection" is in its name. Its sole purpose is to sort students using a normalised standardised scale, helpful for students to position their plans and expectations, and for statisticians for analysing and comparing results. However, a significant load of statistical information on the media can frame aspects of PSU's aggregated results as being either shocking (pathological) or a faithful image of the natural distribution of capacities spread around an ideal mean, when, in fact, this normality is a matter of design.

6.2.3 Hands on: Sorting statistics

Alejandra and I met after the first session to plan and adapt what ought to follow. She indicated that students had not yet studied the normal distribution, and indeed their background in probability is somewhat weak. It could take a couple of hours to develop an understanding of the definition in Equation 6.1 or the skills and language for a computation like in Equation 6.2. We must use the short Session 2 wisely since our goal is to use Session 3 to analyse Figure 6.3.

Students do extract and interpret information from Table 6.1 well-rooted in context, and conversations around Table 6.2 show that they have a grasp on order measures, such as quintiles

Table 6.2: PSU score conversion table for Mathematics, admission 2018. Source: DEMRE.

Corrected score	Percentile	PSU score		Corrected score	Percentile	PSU score
0	1	150		38	86	614
1	1	160		39	86	619
2	1	171		40	87	623
3	1	181		41	88	627
4	1	192		42	89	631
5	1	202		43	90	636
6	1	212		44	90	640
7	1	223		45	91	644
8	1	233		46	92	648
9	2	259		47	92	652
10	3	286		48	93	656
11	5	312		49	93	660
12	8	338		50	94	664
13	11	363		51	94	668
14	16	386		52	95	672
15	21	408		53	95	676
16	26	429		54	95	680
17	32	448		55	96	685
18	38	465		56	96	689
19	43	481		57	97	694
20	49	495		58	97	698
21	53	507		59	97	703
22	57	518		60	98	708
23	60	527		61	98	713
24	63	536		62	98	719
25	66	544		63	98	725
26	68	551		64	99	730
27	70	558		65	99	737
28	72	564		66	99	745
29	74	570		67	99	752
30	76	575		68	100	760
31	77	581		69	100	770
32	79	586		70	100	781
33	80	591		71	100	795
34	81	596		72	100	809
35	82	601		73	100	822
36	84	605		74	100	836
37	85	610		75	100	850

in terms of socioeconomic status and percentiles in test achievement. The table is limited as it is constructed as a function of CS and does not display the 30th percentile necessary to analyse the tautological news as in Equation 6.2.

A tension builds up between exemplarity and pragmatism (see Section 4.2). Working on normal distribution calculations is exemplary for concepts and procedures of the discipline (instrumental) as well as for how statistics are used and shapes their world (critical), but it would require a time frame we do not count on. Working based on an already constructed table of conversion to PSU scores hides its normal distribution grounds, but it does capture the big idea we intend to discuss: PSU scores are on a relative scale.

From there, we decide to base our discussions around Table 6.2, which is constructed on a normal distribution $\mathcal{N}(500, 100)$ but does not require students to dig into normal distribution probability calculations. In order to scaffold the understanding of PSU as a relative scale, we pose the following exercise:

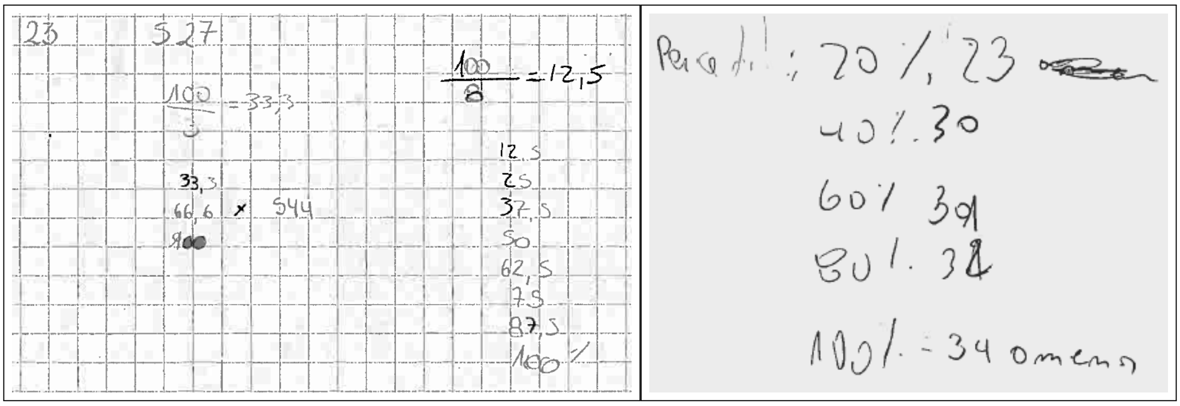
> **Exercise**
> Based on your latest corrected score in a Mathematics PSU practice test, compute your PSU scores in three different ways:
>
> 1. If you had taken the PSU last year
> 2. If the cohort of students was just your group
> 3. If the cohort of students was just your class

The task was successful, and the class results are presented on Table 6.3. See Figure 6.2 as an exemplary output of Andrea and Néstor's work on the task, who belong to different groups. With 23 correct answers, both checked Table 6.2 for the first part of the exercise, yielding 527 points. On the second part of the exercise, however, it is not straight forward, as it depends on the group they belong to:

RAIMUNDO: [*Talking to the group of five students*] But now in your group, in what percentile is each? How would you know what percentile you are in the group?
CLAUDIA: With the "score" (i.e. correct answers) obtained by each of us...
RAIMUNDO: OK, so how do we know who is in the 20th percentile?
NESTOR: The one who got the least [ALEJANDRA *and* RAIMUNDO *nod*].
NESTOR: They are sorted from lowest to highest.

This time, Andrea and Néstor compute differently. On Figure 6.2-(a), Andrea determined her percentile by dividing 100 by 3, the number of students in her group. Her group mates had 22 and 26, respectively, positioning her in the 66.6th percentile, yielding 544 points, as 66th is the closest percentile given in Table 6.2. Néstor, on the other hand, divides 100 by 5 and positions himself in the 20th percentile of his group (see Figure 6.2-(b)), yielding 408 points according to the closest percentile on Table 6.2, the 21st. For this exercise, they had to ask each other's latest corrected score, within the group. Their results are summarised on Table 6.3. It is relevant to locate that and identify how their PSU scores change:

RAIMUNDO: So, what happened to your scores?
CLAUDIA: They went down.



(a) Andrea (b) Néstor

Figure 6.2: Students' calculations of own PSU scores

RAIMUNDO: All of them?

ANDREA: His [*points to Matías*] went down, and the two of us (Andrea and Laura) went up.

MIGUEL: Mine went up.

NESTOR: It went down for me.

RAIMUNDO: It is curious that we are talking about the same number of correct answers, but just because it depends on the rest, it (PSU score) changes.

Then they go for the third part of the exercise, starting by counting themselves to divide 100 in the eight relevant percentiles (see Figure 6.2-(a)) and asking each other's corrected scores. They need to solve the issue of the tie between Andrea and Néstor:

NESTOR: How many did you get, ANDREA?

ANDREA: 23.

NESTOR: I got 23 too.

NESTOR: I stayed almost the same as before. I was on the 20%, and now I am on the 25%.

RAIMUNDO: (...) You (ANDREA) also got 23 [correct answers], right? [ANDREA *nods*] So, is it correct to say "25% of the class got 23 or fewer"?

UNISON: No.

RAIMUNDO: Because, how many had 23 or fewer?

NESTOR: Three!

RAIMUNDO: So what percentile comprises three out of eight?

NESTOR: 25... No! It would be 37.5!

RAIMUNDO: I will ask again. So, would it be correct to say that 37.5% of the class obtained 23 correct answers or fewer?

UNISON: Yes!

Sorted from the lowest to the highest, Néstor sees himself as being in the second place (25th percentile), together with Andrea, which is true. However, I remind them of the definition of percentile we have been using, facilitated with true/false questions, which they answer without much trouble. The dialogue above is followed by independent work, whose results are collected in Table 6.3.

Cohort		2018 Admissions		Group		Class	
Corrected score	Group	Percentile	PSU score	Percentile	PSU score	Percentile	PSU score
22	1	57	518	33.3	448	12.5	363
23	1	60	527	66.6	544	37.5	465
23	2	60	527	20.0	408	37.5	465
26	1	68	551	100.0	850	50.0	495
30	2	76	575	40.0	465	62.5	536
31	2	77	571	60.0	527	75.0	575
32	2	79	596	80.0	591	87.5	627
34	2	81	596	100.0	850	100.0	850

Table 6.3: Students PSU scores computed in three ways

The purpose of our short session was to scaffold the sorting nature of PSU scores, without explicitly addressing its normal-distribution-based construction. Their surprise to see how their scores change according to the cohort let them identify that PSU is not only about personal achievement; it is a competition:

RAIMUNDO: Imagine that only you as a class were taking the PSU test. Or just one group...
DAVID: [*Joking, to* BRUNO] I wish you do bad [on the test] then, so I can do well.

6.3 Session 3: Graphs in political arithmetic

On the Christmas of 2016, Chilean news outlet *La Tercera* published an article titled "Emblematic secondary schools get out of the 50 best schools in PSU ranking" (Mardones, 2016), featuring the diagram on Figure 6.3. It referred to two of the most prestigious public secondary schools in the country and described how their standardised test scores had dropped over the last few years, to the point of not reaching the top 50, as they historically do. In an electoral year's eve, politicians and pundits blamed both the students' and the government for the decline. The former, for participating in strikes and protests that led to educational reforms. The latter, for the reforms themselves established by at-the-time president Michelle Bachelet (2014–2017). The narrative became well-established. One candidate started a campaign called "save the emblematic [schools]"[10], supported by political and academic figures, arguing for the right of some schools to select their students with academic admission tests. The opposition coalition won the presidential election, with Sebastián Piñera at its head. His educational agenda was centred around countering some of the reforms previously implemented, in particular, to allow schools select pupils through admission tests and families to pay for subsidised-schools. All of this, under the motto of *parent's freedom to choose an education for their children*.

The case of Figure 6.3 is exemplary for what is considered to be the *prescriptive* function of mathematics in society, referred to "those situations where mathematics leads to human action or automatically to some sort of technological action" (Davis & Hersh, 1986, p. 120). On the one hand, PSU and its results produce a type of automatic technological action at the core of the mechanism of higher education admissions. It constructs an accepted reality, defining who is

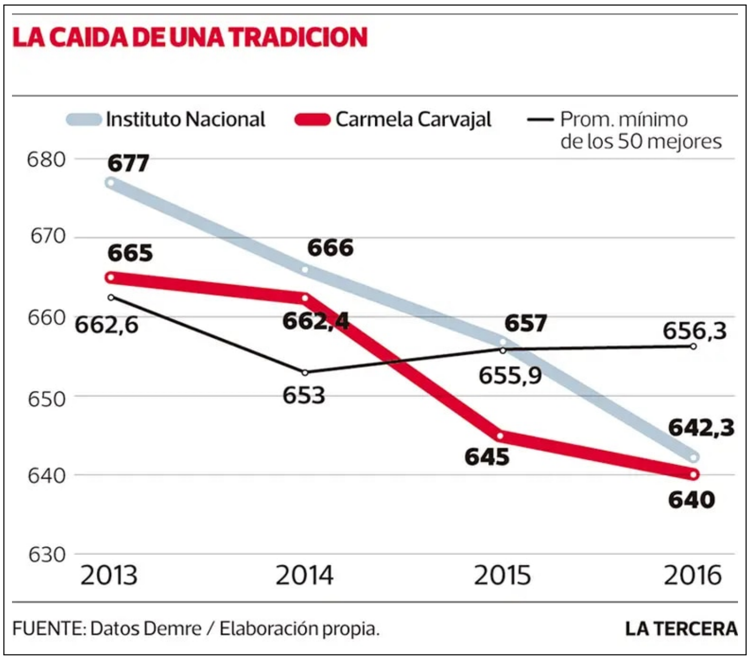


Figure 6.3: The fall of a tradition. Source: La Tercera

eligible to join higher education programmes. On the other hand, the reporting of (selected) PSU score trends is not merely descriptive or predictive either. It leads to human action, prescribing the alleged necessity to change the political course.

Here lies the motivation to discuss political examples in the classroom, "because they clearly and explicitly illustrate the kinds of consequences that may result of the application of mathematics in such contexts" (Sánchez Aguilar & Blomhøj, 2010, p. 20). Students can experience the relevance and societal importance of statistics by working with political contexts. Moreover, these contexts can be memorable anchoring points for students understanding of statistical concepts and ideas and insight in the role and function of mathematics in societies. This awareness is of particular relevance with statistics, which, as pointed out in Chapter 1, has the political at its core as a subject-matter.

6.3.1 Political arithmetic

Just like with the normal distribution, I like to ask my students about the origin of the word *statistics*. The association with the word *state* is immediate among their answers but in the sense of a system's *status*. From their intuition, statistics is believed to be coined as it describes the state of quantitative variables. Statistics would have acquired its name after its methodological nature.

As mentioned in Section 3.1, statistics as the activity of collecting and reporting data can be traced back to censuses in antiquity, but the term statistics did not emerge until the notion of State-nation in Europe rose. *Staatswissenschaft* (Achenwall, 1749) or *statistik*, was coined as State-science, translated to English as political arithmetic, through which numerical data were collected and reported about matters of the State-nations. Political arithmeticians already noticed the political uses of their work. For example, Edmund Halley's tables of births and funerals in Breslaw (Halley, 1693a) was not for the sake of merely collecting facts, but he acknowledges that it is anchored in displaying "the strength and glory of a King in the multitude of his subjects" (Halley, 1693b, p. 656). Early statistics had a political purpose; it was built under the rationale of numbers displaying the State-nation's grandiosity and strength.

Political arithmetic required a political conceptual framework of sovereignty and power, the idea of the State-nation as a unit, different and eventually superior to others. This conceptual framework came after the obsolescence of the feud as the political belonging, and before the modern framework of a population – broader than and autonomous from States – led to social statistics (Donnelly, 1998). In that sense, ideas of addition, distribution and means helped to hold things together (Desrosières, 1991) and configure the State. For example, advancements in epidemiological studies were conceived by assessing one nation or comparing nations with each other, as it was the Anglo-French debate on suicide between Esquirol and Burrows (Hacking, 1990, pp. 64–72). Not only did they involve mortality registries as raw data, but preconceptions on how French and English people are played a significant role in conjecturing about the causes of suicide. As Donnelly puts it, "what made political arithmetic *political* were, in sum, two elements: its not infrequently explicit political aims; and its implicit conceptual frame of reference – the community conceived first and foremost as a political realm as a creation of political will" (1998, p. 232).

These characteristics of political arithmetic can be identified in partisan politics today, under a more local scope. For once, the political aim of constructing, collecting and displaying statistical information is spot on, though often proclaimed as being objective facts. The selection and reporting of statistical messages by one political party will certainly be to confirm and exacerbate its popularity and achievements. On the other hand, it is necessary to delineate who is *us* and who are the *others*; right and left, government and opposition, conservatives and liberals, present and prior administrations.s

6.3.2 Reading graphics

After a recapitulation of the lessons from the exercises in Session 2, Alejandra and I wanted to introduce Session 3 by showing some cases of controversial graphical representations of data. One example that came to my mind is shown on Figure 6.4-(a). During a presidential debate, then-candidate Sebastián Piñera displayed a disproportionate bar graph of crime rates on selected years of three different administrations, including his own previous government at the centre. His message is simple and clean: elect *us*, when the *others* are in power, crime rates rise. I was not sure about using this example, as I did not want to be partisan. Alejandra insisted on its relevance and also suggested to show one she remembered well and considered to be on the same line. Published in the newspaper *El Mercurio*, Figure 6.4-(b) displays a bar graph of poverty rates (defined by family income) at the end of the three latest presidential administrations. The message is similar to the previous one: one coalition's government comes with a decrease in poverty, which increases with the others. The reader will notice that the bars are not only disproportionate, but their relative sizes also do not even match the sorting of numbers they represent.

Our purpose is for students to read and interpret the graphs and notice errors or questionable features from them. In subsection 2.3.3, I have pointed out that in this chapter, the intention is to use a bottom-up framework for analysis. As a result, students insights provide us with a local

(a) Sebastián Piñera on presidential debate (b) Poverty by administration

Figure 6.4: Introductory graphs for Session 3

proto-framework to analyse and construct political graphs. In summary, students raise – more or less spontaneously – three main aspects: (G1) source and interpretation of data points, (G2) meaning and completeness on the X-axis, (G3) position and proportion on the Y-axis, (G4) overall purpose or message conveyed, and (G5) what "people" would see.

Mathematical-oriented reflections

Both graphs specify the sources of data, though they are not quite legible for students. We do mention data sources as they provide context for them to interpret data points as a straight forward exercise of *reading data* and *reading between data* (Curcio, 1987). For example, by looking at Figure 6.4-(a):

RAIMUNDO: Every three years. OK, and what is different about those years?
NESTOR AND CLAUDIA: (. . .) In 2010 it (crime) increased, in 2013 it decreased and in 2016...

The X-axis represents time in our examples, and students question whether there is some cherry-picking. On Figure 6.4-(a), Néstor interprets the time variable on the X-axis and notices that the timespan does not make sense to him:

RAIMUNDO: (. . .) So what do you have there? There are three bars...
NESTOR: Every three years.
NESTOR: Every how many years is the change of president?
RAIMUNDO: Every four (. . .)
NESTOR: But here it is taken every three years...

By looking at Figure 6.4-(b) they even speculate about the reasons for choosing certain years, but a little help is needed:

RAIMUNDO: (. . .) Why must they have chosen those numbers, from 2009, 2013 and 2015?
MIGUEL: To evidence the change?
NESTOR: For marketing.
CLAUDIA: Maybe in 2009 poverty increased too much, much more than in 2006.

RAIMUNDO: (...) When a government just started, one cannot demand much. However, by the end, in 2009, you measure it (poverty). Here, in 2013. But here...

NESTOR: 2017.

MIGUEL: Maybe it was not convenient (...) it was not convenient to show 2017.

The most prominent issue to focus on both examples is the proportionality of the bars against the numerical values they represent. It is an assessment of the variable on the Y-axis, which does not require much prompting to be noticed. On Figure 6.4-(a):

NESTOR AND CLAUDIA: (...) In 2010 it (crime) increased, in 2013 it decreased and in 2016...

MIGUEL: It is disproportionate! That 22.8% cannot be there!

RAIMUNDO: Why? That is interesting.

MIGUEL: Because it is low.

RAIMUNDO: (...) What does disproportion mean?

MIGUEL: That, for example, 30.7[%] is too tall in comparison. And 22.8[%] is shown as being lower (...) it is not even half [of the bar size] of 30.7[%]. And half of 30 should be 15, and the blue one is beyond 15.

On Figure 6.4-(b) the disproportion is more grotesque, so it is noticed immediately:

RAIMUNDO: Let us go to the next one [*Figure 6.4-(b) shows up on the screen*]. Your teacher showed me this one yesterday...

MIGUEL: That 11% is wrong there!

RAIMUNDO: What is it wrong about it?

BRUNO: It should be shorter...

LAURA: It should be shorter than the [bar with] 14.4[%].

Reading information and assessing both axes (G1, G2, G3) requires students to understand what variable each axis represents. However, as much as these utterances are rooted in context, they can be exemplary for what Skovsmose refers to as mathematics-oriented reflections (Skovsmose, 1998; Barbosa, 2006). Reading data points or eventually checking sources is a right/wrong question of the type "are we interpreting the data point correctly?", the first two entry points on reflective knowing (Skovsmose, 1992, 1994). Completeness on the X-axis and positions on the Y-axis are associated to the question "does the graph have the correct information?" That is, whether the information is correctly and wholly represented.

Model-oriented reflections

Within statistics, context gives meaning to variables and their interaction. Furthermore, the political context brings students to interpret the overall message that is conveyed through the graph. It is past a mere description or evaluation of the graph's construction. For example, on Figure 6.4-(a):

RAIMUNDO: In what context was that? In what moment did he show that graph?

MIGUEL: When... before he became president?

RAIMUNDO: Yes. Anyone else remembers? Did you see it on television?

NESTOR: No.

RAIMUNDO: It was during a debate, OK?

ALEJANDRA: Come on, guys! How do you not remember!
RAIMUNDO: (...) Imagine, on a presidential debate, where everyone is watching and thinking about who will they vote for, why is it important for him to show this graph? What did he want to say?
NESTOR: That he will (...) reduce crime.
MIGUEL: That crime decreased during his period.

Néstor and Miguel have a slightly different interpretation. Using Curcio's (1987) typology, Néstor is *reading beyond data* (projecting an eventual next administration), while Miguel is *reading between data* (from bar to bar). What both answers have in common is the recognition of the political message: one president or one coalition is better dealing with crime than the others. As much as the X-axis represents time in years, political arithmetic makes use of the conceptual framework of *us* and *others*. In the example, those years represent political coalitions, not time. More importantly, the selection of years to report crime rates is for the sole purpose of display (we can deal with criminality) and maintenance (elect us) of power.

Students ability to see the intention beyond the graph is also evident about Figure 6.4-(b):

RAIMUNDO: (...) If we just look at the graph, who finds it convenient?
NESTOR AND LAURA: To Piñera.
RAIMUNDO: Why?
LAURA: Because it is the lowest...
BRUNO: Because the graph shows a 14[%], but the graph is like this [*hand gesture of convex function*], lower (...) It is saying that Bachelet...
MIGUEL: It is saying that with Bachelet poverty increased. With him (Piñera), it did not.

This feature of acknowledging the political intention is beyond reading data or projecting it. The graphical representation acquires a human dimension and an implicit claim. It can be positioned as technological knowing, evident by model-oriented reflections (Skovsmose, 1998), since students are not merely asking about the correctness of the graphical representation, but its fit to a particular purpose. Perhaps the following comment makes the human aspect more clear:

RAIMUNDO: (...) The source of information – or, as you (Miguel) said, the percentages we see there – come from CASEN[11]. But the graph...
ANDREA: ... is made by somebody else.

Context-oriented reflections

The final aspect brought up by students relates to how the graphs affect people's conception of an issue, be it crime or poverty. For example, after analysing the proportions on Figure 6.4-(a), the following episode takes place spontaneously:

NESTOR: It is just that he did it for people to see it barely. There are people who do not evaluate the numbers
LAURA: People go by the size that decreases and not by the numbers they represent.
MIGUEL: The idea is for them only to see that it (crime) decreased.

The same reaction occurred after Figure 6.4-(b):

LAURA: People just go by the size of the graphs. A lot of the time, they do not mind the numbers over there.

This aspect (G4) resonates with the fifth entry point on reflective knowing (Skovsmose, 1994), namely "how does this application affect our conception of a part of the world?" Students are reflecting beyond the intention of the author of the graph. They are acknowledging that how the graph is made affects our conception of how one or other political coalition has an incidence on the country's crime and poverty. Indeed, the reflection is triggered by the graph's wrongness, but awareness is taking place on what it produces in society.

Testing the proto-framework

We put this five-aspect proto-framework to the test, by finally addressing Figure 6.3. Students require some contextualisation, such as whom these two schools are and why they are in the graph, or what the black line means (top 50 schools). This contextualisation is solved mostly amongst them.

The source of information is stated below the graph, and I give it away to make the point that the information is most certainly true. From that point, students interpret the graph quickly (G1):

NESTOR: It is representing that the one who stays the most... oh no...
BRUNO: It represents where the best academic excellence lies, all of that...
NESTOR: And how they have been decreasing.
CLAUDIA: How they have been decreasing until 2016.
NESTOR: From 2013 to 2016.

Students recognise which data are selected: so-called emblematic schools, somehow staying at the top, yet decreasing in their PSU scores over the years. The distinction between Claudia's and Néstor's answers will become relevant for asking questions about the X-axis. Nevertheless, the political intention (G4) is addressed before, after clarifying who were presidents over the years:

NESTOR: There they are saying that Bachelet made... or her administration...
MIGUEL: ... caused the decrease...
...
CLAUDIA: They are making her responsible for the drop.
BRUNO: There it is. In 2013 it was the best and then when Bachelet assumed, it began to drop. And now it says that if Piñera takes over again, it will rise.

Just as in the introductory examples, the political message takes both the form of *reading between data* (the trend under the current government) and *reading beyond data* (what would happen if a new coalition governs in the future) (Curcio, 1987). Just as in political arithmetic, this interpretation needs the more or less explicit grouping of years into political coalitions.

Three aspects of the Y-axis are questioned (G3). The first concerns the relative position of data against the axis, as a simple right/wrong issue:

RAIMUNDO: How do you know that [data are well represented]?
LAURA: Because of the numbers that are to the side [*vertical hand gesture*].

RAIMUNDO: (...) Let us see. For example, 677, is it OK?

CLAUDIA: [*Checks on the Y-axis*] Yes.

RAIMUNDO: OK. 666, 657, 665...

NESTOR: [*Checks on the Y-axis*]. They are all fine.

Second, they go beyond the reading of data to question how significant are the changes represented:

RAIMUNDO: (...) Do you think the decrease is significant?

MIGUEL AND NESTOR: Yes.

RAIMUNDO: Yes? Why? Where do you see that the decrease is significant?

NESTOR: From... let me see, six hundred...

MIGUEL: By the path they follow...

NESTOR: By the path they follow it is important, but the numbers drop just by 40.

CLAUDIA: So it does not decrease so much.

The third aspect concerns the axis itself, as their previous knowledge on the PSU scale is activated:

NESTOR: (...) The scale [*referring to the Y-axis*] goes from 630 to 680, that is why the difference is quite remarkable.

RAIMUNDO: Oh, because we are only starting from here [*points to the origin of the axis*]...

NESTOR: Yes. 630 to 680. Because if it were from 0, it would not be...

MIGUEL: It would not be so noticeable.

...

NESTOR: Because 0 would be the minimum [PSU] score.

MIGUEL: Wait, not really!

RAIMUNDO: What is the minimum score?

NESTOR: No, wait... It is 150!

At this point, the discussion with students has let us – as a class – find issues to focus on when reading Figure 6.3, elicited by mathematical and technological reflections (Barbosa, 2006). In our self-made proto-framework, adherence to the source (G1) and positions on the Y-axis (G3) are evidence of mathematical knowing, while the recognition of the purpose behind the graph (G4) is evidence of technological knowing (Skovsmose, 1994). The reader will notice that students' reflection on how the graph shapes of perceptions on the introductory examples are not illustrated while analysing *the fall of a tradition*, which would entail evidence for critical reflections. Yet.

Students begin to show unease about the reach of the Y-axis in terms of the PSU scale. As much as they do find questionable issues about the graphical representation acknowledging the political framing, alternative versions of the graph are still anyone's guess. With Alejandra, we notice that it is time for them to construct their own graphs.

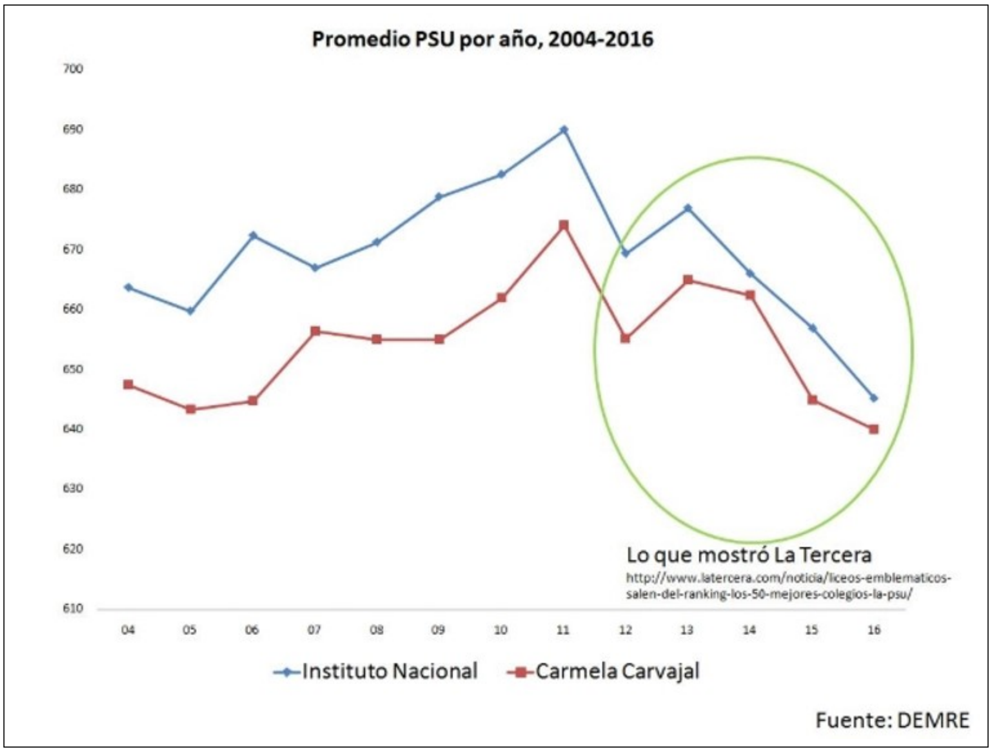


Figure 6.5: PSU trends since 2004. Source: Fábrega (2017)

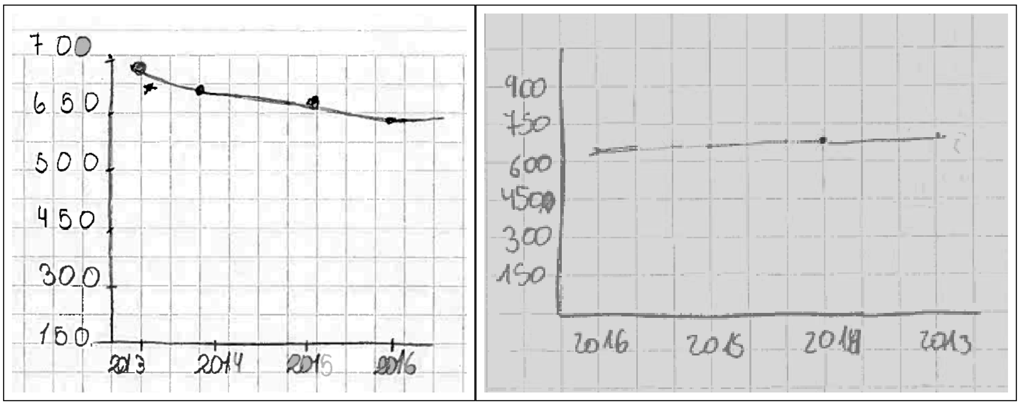
6.3.3 Writing graphics

Two weeks after the publication of the article featuring Figure 6.3 on *La Tercera*, the medium CIPER[12] published an article denouncing many of the issues Alejandra's students spotted (Fábrega, 2017). The article featured a graph including PSU scores for the same two high schools in question, but starting in 2004, the first year PSU was implemented (see Figure 6.5). Just with that visual aid, one can spot that the decrease had already started since 2012, addressing the problem completeness of the X-axis (G2). However, issues with the Y-axis are addressed in other parts of the article. What is worth highlighting is the very fact of making a graphical representation to be critical about someone else's. Making an own *production* of statistical information to be a responsible *consumer* of statistical information (Gal, 2002). There lies an opportunity to write the world with statistics (Gutstein, 2006; Weiland, 2017).

RAIMUNDO: All right. So I will ask each of you to pick one [of the two] high schools, with the same data, and show how it would look like if it (Y-axis) starts at the minimum, which is...
LAURA: 150.

Two of their productions are shown in Figure 6.6, authored by Claudia (a) and Andrea (b), respectively, displaying a much milder decrease in PSU scores, exaggerated in the case of Andrea. Claudia's version has 150 points at the very bottom of the Y-axis but fails to make consistent, jumping from 450 to 500. Andrea succeeds with the basic 150-long jumps, but she is implicitly locating 0 at the bottom of the Y-axis. Besides, Andrea constructs the X-axis at a counter-intuitive orientation. Both are omitting the axes labels as well.

Nevertheless, their conversations during and after the process of constructing their own graphs become more interesting. As soon as they began to work, they take a closer look at the



(a) Claudia (b) Andrea

Figure 6.6: Claudia and Andrea’s version of PSU scores trend

data points:

MIGUEL: Instituto Nacional drops only by 35 points.
RAIMUNDO: From its highest to its lowest?
MIGUEL: Yes.
...
NESTOR: Carmela Carvajal drops by 25 points.
BRUNO: 24.7, actually.
NESTOR: It is the one staying the most... it is not so much of a decrease.
MIGUEL: 25.3.
BRUNO: [25].5.

In order to produce their own graphical representations, they need to be more precise about the Y-position of data, triggering them to question the differences between one year to another and throughout the timespan 2013-2016. Consequently, the Y-axis component of our proto-framework (G3) is refined by producing their own graph.

Further refinement of G3 comes with reactivating the relative nature of the PSU scale, which they bring up while discussing the political intention (G4) of the graph:

MIGUEL: They want to say that Bachelet does not support... *those* schools. Instead, she supports others more.
...
RAIMUNDO: But there you are saying something interesting. OK, even if we agree upon those two schools decreasing, the other schools...
LAURA: Decrea– I mean, increased!
MIGUEL: Increased!
RAIMUNDO: Do you remember the exercise we...? Let us wait for CLAUDIA, because this is important (...)

Claudia had gone to the bathroom. I wanted to wait for her to make this crucial connection in the plenum. However, the topic of conversation switches to yet another missing aspect of our proto-framework, namely the completeness on the X-axis (G2):

RAIMUNDO: So, how was Instituto Nacional doing before?

NESTOR: Good.

BRUNO: Good, naturally.

LAURA: It prompts us to think it was doing well, but... we do not know.

NESTOR: In 2013... ah, we do not actually know! I mean, in 2013... we do not have the datum for 2012.

When reading the original graph, the assumption that Instituto Nacional and Carmela Carvajal high schools were "doing well" until 2013 was sustained and unquestioned. Now, without additional interventions, they do begin to question this assumption and realise that data before 2013 are missing. My interpretation is that, since they have been constructing their versions of the graph, it is upon their pencils the power to write relevant positions on the axes. The Y-axis (G3) was already addressed by setting the origin to 150 points. The X-axis' (G2) turn came next.

Claudia came back from the bathroom, so it was time to reconnect their exercises on the previous session to the meaning of the PSU scores trends. First, we engage in a brief reminder:

RAIMUNDO: (...) Do you remember how to compute the scores? What happened during the last exercise with your scores?

NESTOR: We used percentiles and number of correct answers.

CLAUDIA: But only with the students we are here.

...

RAIMUNDO: (...) So you computed your scores in three different ways, considering the same number of correct answers.

NESTOR: Some [scores] went up, some others went down.

RAIMUNDO: And why did they change? That is weird if you have the same corrected score.

BRUNO: It depends on the people, the number of people.

MIGUEL: The total people taking the test changed.

After revising some of their results from Table 6.3 we move back to Figure 6.3 to make the connection. Some clarification is needed since, from their experience, it appeared to be that the number of people was the factor affecting their PSU scores.

RAIMUNDO: We have made some criticism to the graph (Figure 6.3), we have done some analyses. However, there is yet another particular element about how the scores are computed. The fact that they (the two schools) have increased or decreased depends only on them?

NESTOR: It depends on how many took the test that year.

RAIMUNDO: And how did they go.

NESTOR: Yes.

RAIMUNDO: So, if a school's PSU score decreases, does it necessarily mean that they had fewer correct answers?

BRUNO: No.

RAIMUNDO: What could have happened? What happened to some of you?

NESTOR: There were more or fewer...

MIGUEL: Maybe there was a rise in participation in PSU.

...

RAIMUNDO: What if you do as well as those schools do this year. What will happen to these two schools' scores?

MIGUEL: They will drop.

NESTOR: Their percentile will drop!

RAIMUNDO: That is it! They will fall into a lower percentile. (...) So is there a nation-wise problem here if those two schools drop?

NESTOR: Because they focus more onto them than others...

CLAUDIA: They always focus more on them.

MIGUEL: Perhaps it could be the other way around, then. We have grown [as a country] because each year there are more good scores.

RAIMUNDO: (...) He is saying that those schools' decrease is even good news? Really? How come?

MIGUEL: Because maybe it is not a problem with those schools. Perhaps they stay the same, but there are other people doing better.

NESTOR: If they decrease, they lower their percentile and others go up.

The dialogue above suggests that the activity during Session 2 made some of them believe that the size of the cohort is the main factor altering their PSU scores, disregarding the sorting. By identifying that issue, I challenge them with a hypothetical question to shift the conversation until Néstor remembers the role of percentiles. Moreover, Miguel realises that a drop in PSU scores is merely a drop in order, not in absolute performance, to the level of regarding the decreasing trend of the two high schools Instituto Nacional and Carmela Carvajal as good news.

Finally, students do engage in a reflective discussion by questioning how this application shapes people's perception (G5):

RAIMUNDO: What does the person who made this graph want to emphasise?

NESTOR: They want to emphasise that [*yawns*]...

MIGUEL: They want the change to be noted. Although, in reality, it was not so much.

NESTOR: I mean, if people saw it like this, without the numbers, they would say: "it was a lot".

NESTOR: I, at least when I first saw it, I said "wow, it decreased a lot", but then I saw the numbers and (...) When one sees things without analysing, one sees... one sees a lot [*General laugh*]

MIGUEL: He is saying the same thing all over again.

In Table 6.4 I roughly summarise the local proto-framework for analysing political graphs developed during the third session and its relation to mathematical, technological and reflective knowing. On the third and fourth column, I mark whether these aspects were evoked by reading and writing graphs based on Figure 6.3. My message is not that students are not able to read political graphs without writing them since the proto-framework was made from them reading graphs during the introduction. However, by changing to the context of PSU, the five aspects of the proto-framework are more easily evidenced by students reflections when writing an alternative version of a political graph.

Moreover, Miguel's realisation that the graph could have an opposite interpretation can be considered yet another element (G6), which I named *alternative interpretation*. The sixth entry point on reflective knowing (Skovsmose, 1994) is evidenced by the question "could have we reflected on a different way?" This point goes beyond the realisation of the political message (G4) and the realisation on how the graphical representation affects our perception (G5), but it means that students are even suggesting alternative perspectives to the one given by the graphical production.

Code	Graph aspect	Reflections	Reading	Writing
G1	Correct data points	Mathematical	×	×
G2	Completeness on X-axis			×
G3	Position/proportion on Y-axis		×	×
G4	Political message	Technological	×	×
G5	Perception shaping	Critical		×
G6	Alternative interpretation			×

Table 6.4: Proto-framework for PSU graphs

As a topic of research, Monteiro and Ainley (2007) argue that the interpretation of graphs is at large focused on cognitive skills (Curcio, 1987; Friel et al., 2001). Instead, they offer a perspective that involves a *critical sense* "to encapsulate the way in which a sophisticated *reading* of graphs involves mobilising a range of different kinds of knowledge and experience" (Monteiro & Ainley, 2007, p. 202). Their point is to balance the different aspects, including data collection, social context and personal experiences. In my view, the above episodes show that such a critical aspect is enriched by balancing different aspects when *writing* alternative graphs.

6.4 Moving forward

This chapter shows the potential of a learning environment to engage students in critical reflections about the formatting power of statistics, through a case that is close to their experience. Before moving forward to the next chapter, it is worth to call attention to some learned lessons in light of the design principles.

6.4.1 Exemplarity, available or not?

The case of PSU results depicted on the media can be *exemplary* from all three perspectives considered. First, it draws on students' personal experiences, as they are well informed – sometimes better than I am – about the underpinnings of the tests and its consequences. Moreover, they use their context knowledge about the political framing of statistical graphs.

Second, it is exemplary for some of the big disciplinary ideas, namely the idea of distribution. Garfield and Ben-Zvi (2008) suggest that two main types of distributions are experienced by students: empirical and theoretical. Students explore the empirical distribution of corrected scores in Session 2 by arranging them in order and computing PSU scores under different cohort scenarios. The normal distribution plays a particular role since usually theoretical distributions are connected to empirical ones as a way of contrast and modelling (Wild, 2006), while here the normal distribution is *imposed* on the empirical one via Equation 6.1. However, the normal distribution is not made available to students. This pedagogical decision was taken with Alejandra, on the *pragmatic* grounds of time availability.

Throughout the chapter, I have made some connections between PSU and historical developments of statistics, such as social statistics, political arithmetic, as well as technicalities that produce tautological news (Section 6.2). These connections are rooted in the use of the normal distribution as an ideal or as a practical tool. As a researcher, I can experience PSU’s exemplarity in the process of analysing and writing this document, but these reflections are obscured to the teacher and students. Making the normal distribution available to students could be an opportunity to make PSU and related events exemplary not only as a quality of the case but as a quality of the teaching-learning practice.

Third, the learning environment shows great potential concerning the use of statistics as a tool not only to describe the world or to make predictions but to prescribe technological or human action (Davis & Hersh, 1986). On the one hand, PSU is not just a battery of tests to inform students about their academic performance. Results take part in a unique nationwide technological system of admissions to higher education that is undisputed and automatic, and thus prescribing access the distribution of students throughout higher education institutions. As David puts it, “in the end, PSU measures your future”. On the other hand, the case explored in Session 3 is exemplary of statistical information to prescribe deliberate human action, namely the alleged need to change political course.

Regarding the latter, using a local proto-framework for analysing political graphs, students’ critical reflections reveal an awareness of the formatting power of mathematics (G5) and propose alternative systemic interpretations (G6) to the one conveyed by a graph as a political technology. Néstor insists on how our perception is affected by a decreasing graph if not scrutinised (G5). Moreover, Miguel makes use of the relative character of PSU scores and conjectures that “perhaps it could be the other way around, then. We have grown [as a country] because each year there are more good scores” (G6).

Part of the *problématique* of the thesis involves asking to what extent can or should the preoccupations of CME take place in the classroom discussions. In Session 1, students make connections from the technical term “percentile” – necessary to move onto the inquiry – to the use of quintiles to characterise their socioeconomic background as belonging to specific groups of the population. After David sharing his after-school prospects, I have speculated a connection to the notion of foregrounds, as part of the dialectic between structure and individual (Wedege, 2016). David’s preference to join the military is expressed in terms of individual preference (“I do not like studying”), yet rooted in structural issues, like the fact that for the armed forces “even healthcare is for free”. PSU results reflect some systemic inequalities in school performance, as the introductory news article shows (Tele 13, 2017). Making these reflections available to students may seem compelling and adequate from a CME perspective on social justice. However, if students become aware of these structural inequalities, it can perpetuate a sense of fatalism that would go against ideals of inclusion. This risk is pinpointed by Valero et al. (2015) about making a separate curriculum for the “losers”, inspired in ideals of social justice. Labelling different groups of people, for the sake of inclusion, can produce a double gesture of exclusion (Popkewitz & Lindblad, 2018).o

Therefore, a tension is revealed in the notion of exemplarity, and it relates to the distinction between analytic and normative approaches to problems of research (Niss, 1993). On the one end, it seems self-understood that making mathematical notions available for the students can make the case instrumentally exemplary for the use of statistical distributions. The question is, therefore, *how* to do so and suggests an analytic approach. Whereas, making available the systemic inequalities and structural forces that influence students' foregrounds is an ethical dilemma. The question is, therefore, *why* to do so and suggests a normative approach.

6.4.2 Inquiry approach

One of the lessons learned in the previous chapter was that of not letting the establishment of necessary mathematical notions to the final part of the inquiry work. The work done on Session 2, where students took the time to compute PSU scores, was a necessary step to inform their reflections on Session 3, echoing the relevance of constant support and scaffolding in inquiry-based approaches to teaching and learning (Hmelo-Silver et al., 2007; Blomhøj, 2016)

The examination of a political graph on Session 3 had an interesting shift regarding students' independent work: constructing their version of a graphical representation enabled further critique on the representation they were asked to evaluate. It confirms the suspicion settled about the "Guilty or innocent?" learning environment (Chapter 5), wherein the passive role of the students did not enable critique in the classroom. Evidence suggests that letting students be *producers* of statistical information (Weiland, 2017) improves the refinement of the proto-framework, making them better *consumers* of statistical information. When they construct their own graphs, all the same categories of analysis arise, and some are added up.

6.4.3 Pragmatism in collaboration

The principle of *pragmatism* played a role in adapting the sequence in a way it could be approachable for the students within the time frame. In particular, we decided not to teach the normal distribution, as it was not necessary to explore the nature of the PSU scale as a sorting device.

The process of designing and teaching in collaboration with Alejandra conveyed some lessons to build on. Pragmatism should take the form of collaborative *practical organisation* in the form of constant negotiation (Skovsmose & Borba, 2004). Though several ideas and adaptations to the task were negotiated and planned, Alejandra's participation in the project was initiated by my relation to the school's headmasters. It may explain why she did not follow up on my request to discuss the experience in retrospect. As a consequence, the interpretation of students' reflections is exclusively mine. It will be different in the next chapters.

Notes

[8]In Spanish, the words *quintil* and *quinto/quinta* mean, respectively, quintile and fifth.

[9]Until the admission process of 2012, as a disincentive to answering by guessing, incorrect answers used to be punished by discounting one full correct answer for every four incorrect, thus *correcting* the score before computing the standard or PSU score.

[10]Used to be available on http://www.salvemoslosemblematicos.cl

[11]CASEN *Caracterización Socioeconómica Nacional* is the biannual National Socioeconomic Characterisation survey.

[12]CIPER (*Centro de Investigación Periodística*) stands for Centre for Journalistic Investigation.

Chapter 7

“Should we install a thermoelectric power plant?”

RAIMUNDO: ... One is either not allowed to live there or bound to deal with a poor quality of life, but we accept it because we have to put industries somewhere.
BERNARDO: The thing about a sacrifice zone... The idea is to declare it *before* people already live there.

In 2015, the school La Greda, of Quintero, a Chilean central littoral city, closed its doors after 40 students were confirmed intoxicated from polluting agents (TVN, 2018). This incident has not been the first nor the last pollution-related story in the province of Quintero and Puchuncaví, host of an ever-controversial industrial park.

Bernardo is a mathematics teacher at a school belonging to the province of Marga Marga, neighbouring Quintero and Puchuncaví. No clear preference over an inquiry was agreed upon during our first exchanges, until August 2018, when yet another atmospheric saturation event had taken place (Fajardo, 2018). The topic is highly sensitive in the country at the moment, let alone in the Valparaíso region, host of both provinces.

For my research purposes, this event is illustrative of many issues of the risk society. Among others, it shows how some technologies become the source of new risks, i.e. the phenomenon of manufactured uncertainty; and it embraces the complexity of making decisions under risk, both concerning “objective” calculations (probabilities and impacts) and subjective considerations (how people want to live) (Beck, 2000, p. 221). The mathematised morality problematised by Beck (e.g. 1992, 2000) is rooted in probability calculations. These calculations often include multiplications of probabilities of different events that must fail together to produce a catastrophic event, but “independence of the failure behaviour of the redundant items is a mere assumption” (Borovcnik, 2017, p. 1502). This issue was, for example, part of the Danish-Swedish debate regarding the Barsebäck nuclear plant (Gjørup et al., 1982; Löfstedt, 1996), as well as in the aftermath of the Fukushima nuclear plant accident, where all systems failed at once caused by a single input; the tsunami. For Bernardo’s teaching purposes, the inquiry could be positioned in a particular week which he had assigned to cover conditional probability, independence and the multiplicative principle.

Bernardo had told me about an ongoing project for a new fossil-fueled power plant. This time the plant was to be installed in their province, leading to a resistance movement within the community. It seemed appropriate to tackle the issue by contextualising our inquiry in the real event, but the rest of the problem-posing is fictitious; the company's name and risk statements, and available data to work. This choice is due to many reasons. First, the project was too young for public data to be available, as the environmental evaluation was still a work in progress. Second, even if it were available, the technicalities in the evaluation might overwhelm the students' context and academic knowledge base. Let us face it: ours as well. Finally, we did not want to find ourselves in a defamation controversy with the company. We were completely transparent with the students about the fictitious character of the inquiry.

The contingency has the potential of being turned into a learning environment compatible with all three kinds of exemplarity. It is close to the students' personal experience as neighbours of a recent environmental disaster and eventually hosts of a new power plant. It is critically exemplary of several issues of the risk society. It is instrumental to probabilistic ideas, assumptions and calculations as well.

The overall inquiry is simple: should we install the power plant or not? We organise it into a didactical sequence in three 90 minute sessions available:

1. Setting the scene by recalling the disaster taking place in Quintero and Puchuncaví, and informing the new project in their province. Evaluating the probability calculation of an environmental disaster to happen on any day, as given by the fictitious company. Producing students' own calculations after appropriate corrections.

2. Expanding of the previous calculation onto a longer time span.

3. The making and grounding of the decision embedded in the general inquiry; should we install the power plant or not.

In accordance to the general *problématique*, I am addressing the justification for teaching probability "for all" in high school, which falls into the notion of probability literacy, as first conceived by Gal (2005) assuming citizens to be consumers and not producers of probabilistic messages. However, Gal acknowledges that "probability literacy is sometimes called for in generative situations or subsumed as part of both personal and collective decision-making processes" (Gal, 2005, p. 55), though without references to this claim.

An inquiry in the form of decision making under a risk scenario necessitates a more specific and action-oriented framework. The state of the art is that researchers "perceive probabilistic literacy as the ability to use relevant concepts and methods in everyday context and problems" (Borovcnik, 2017, p. 1500). These central ideas can be summarised from Batanero and Borovcnik (2016) into three categories (Borovcnik, 2017, pp. 1500-1501): (1) the theoretical character of probability (including meanings of probability, independence and the problem of small probabilities), (2) conditional probabilities (dependence on prior judgements and asymmetry), and (3) concepts building on probabilistic evidence (e.g. correlation). These can be interpreted as the main themes to which one expects students' conversations refer.

On the other hand, within these themes, Borovcnik (2011) defines five fundamental abilities of probabilistic thinking, which he later summarises into four in the context of risk (Borovcnik, 2015, 2017): (1) to balance psychological and formal elements, (2) to understand that direct criteria for success are nonexistent, (3) to separate between randomness and causality, and (4) to separate reflecting and deciding upon a situation.

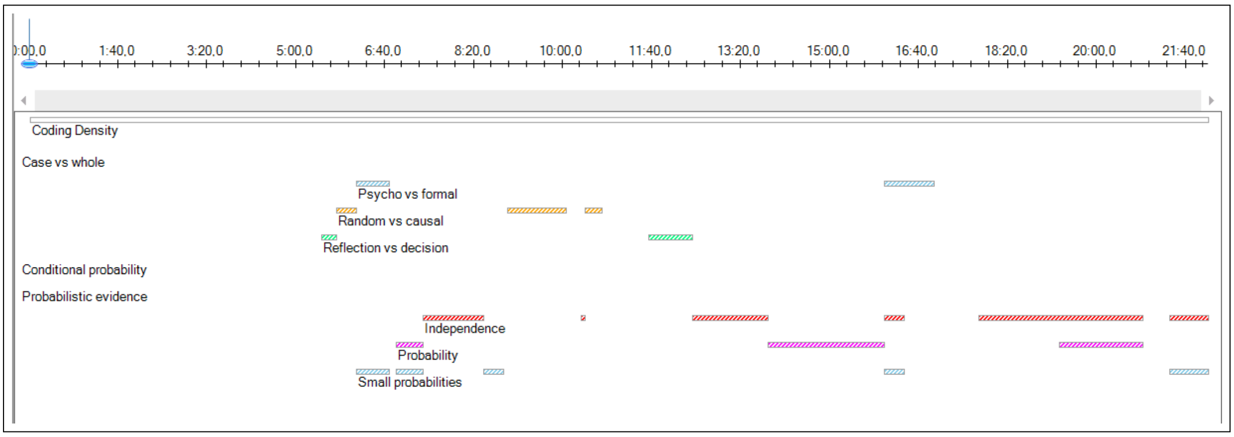
The analysis is made about students' comments during the joint reflection taking place during Session 3, by coding students' utterances according to both frameworks, i.e. by the three main themes or aspects of probabilistic literacy and the four central abilities of probability thinking. As a result, the coding can reveal what relevant themes (literacy) are students reflecting upon (thinking). An overview of the coding can be seen in Figure 7.1, as depicted in the software NVivo. The first category "theoretical character of probability" has a degree of coverage of 47% and 37% in the first and second half of the joint reflection, respectively. Thus, it is more usefully displayed divided into its three subcategories (probability, independence and small probabilities).

An aspect worth highlighting from this general overview is that the probability literacy aspect of "conditional probabilities" is barely not addressed, and "probabilistic evidence" is not found at all. This coding is yet another reason to unfold the "theoretical character of probability" category. In fact, the only utterance seen in Figure 7.1-(b) associated to conditional probabilities is mine, attempting to reinterpret a student's comment, which to a certain extent could have been coded as being a discussion about independence. This absence is not surprising, considering the context of the learning environment, which does not tackle – at least, not explicitly – the subcategories within those two categories: asymmetry of conditional probabilities, the judgement of prior probabilities employing the Bayes' formula (conditional probability) (Borovcnik, 2012), and correlation and relative odds (probabilistic evidence) (Borovcnik, 2017).

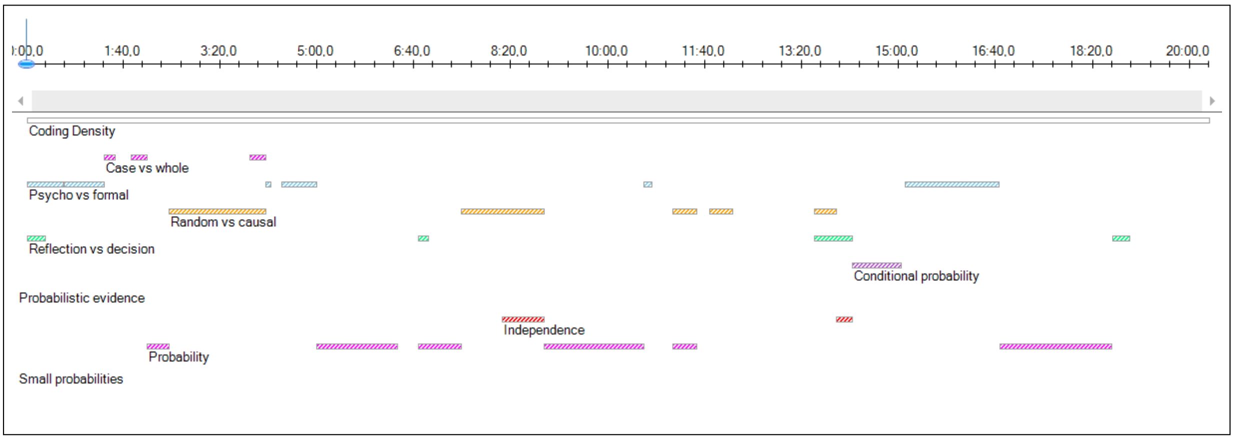
At this stage of the analysis, the research question about the critical justification for probability is not yet addressed. I intend to describe whether and how critique is embedded in students' discussions, in the sense of reflective knowing (Skovsmose, 1994; Barbosa, 2006), connecting it to the topics and abilities coded in the analysis of transcripts o the joint reflection (Session 3), as discussed in Section 4.3. However, some relevant ideas emerge during the whole experience. As a frame of reference, I make use of the conceptualisation of the risk society given by Beck (2000) to position these emerging topics on a critical level.

Most of the transcriptions displayed in this chapter refer to the last session, where one student per group is designated to share their conclusions. Therefore, a few of the participating students are given names in the writing of this chapter. The classroom experience is carried out with Bernardo's 11th grade class of 25 students, whom we arrange into five groups of five students each:

- Sergio, Sandra and three more
- Manuel and two more
- Cristian, Martina, Gaspar and two more
- Pedro and four more
- Nicolás and four more



(a) First half of joint reflection



(b) Second half of joint reflection

Figure 7.1: Coding according to probability literacy and thinking (Borovcnik, 2017)

7.1 Session 1: The probability of a disaster

Session 1 has three purposes: to set the scene for the context of the guiding inquiry, to evaluate the company's probability calculation of an environmental disaster, and for students to make corrections and their own probability calculation for the same event.

7.1.1 Risk and uncertainty

Risk is a term used with many connotations (Hansson, 2009; Borovcnik, 2015). For some, it is close to a hazard, i.e. a possible negative outcome that should be avoided at all costs. It could mean a positive opportunity, much in the way financial risk assessment proffers monetary value to risk-taking. What is somehow commonplace is that statements about risks involve both the likelihood of an event and its possible impact. Whether the event will happen is unknown. Whether the components of risks can be mathematised, it depends. The historical development of probability is entangled in this tension between what can and cannot be calculated, as Hacking (1975, p. 12) puts it:

> It is notable that the probability that emerged so suddenly is Janus-faced. On the one side, it is statistical, concerning itself with stochastic laws of chance processes. On the other side, it is epistemological, dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background.

As probability became an essential tool for economic theory (e.g. Keynes, 1973; Von Neumann & Morgenstern, 1944; Friedman & Savage, 1948), this duality is recognised to make a distinction of terms between risk and uncertainty, in order to split the waters of what can be approached through mathematical models and what cannot:

> The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; (...) It will appear that a measurable uncertainty, or 'risk' proper (...) is so far different from an immeasurable one that it is not in effect an uncertainty at all. We shall accordingly restrict the term 'uncertainty' to cases of the non-quantitative type.
>
> (Knight (1921, p. 19) as cited in Borovcnik (2015))

It is, therefore, no surprise that this duality is highlighted by Beck's characterisation of the global risk society, where manufactured uncertainties (e.g. a new power plant implies new risks) synthesise the notions of knowledge and unawareness. Two meanings are conflated, "namely, risk assessment based on empirical knowledge (...) on the one hand, and making decisions and acting on risk in indefinite uncertainty, that is, indeterminacy"(Beck, 2000, p. 216). Though different meanings are attributed to risk and uncertainty, in the risk society, one single phenomenon involves both what can be mathematised and what cannot.

As a point of departure for our activities, students read selected extracts from a news report on the latest environmental disaster occurred in Quintero (TVN, 2018), in order to set the scene. More specifically, the report is about a decontamination plan that failed to be approved and put into practice. One of the paragraphs reads:

> *"We do not know what they are emitting because it is not being measured"*, states Diana Pey, representative of the Medical Collegium of Valparaíso, when referring to the lack of tools for evaluating polluting agents. In this report, *we will not know what are the polluters in the environment*, but we will inform on what has happened or stopped happening within the factors affecting the environmental policy.
>
> (TVN, 2018, emphasis added)

Besides the environmental saturation and its health-related consequences in the population, one of the problems in the case of Quintero and Puchuncaví was the uncertainty about what were exactly the contaminants and which industries were responsible for it. At one point during the joint reflection (Session 3), students bring the case of Quintero and Puchuncaví to question the meaning of the probability of a disaster, which I discuss in Section 7.3. I attempt to evoke the risk-uncertainty tension as a response:

RAIMUNDO: (...) At this point, and it has been a while, do we know what the problem was exactly? What went wrong in Quintero and Puchuncaví?

SERGIO: At least what I think is that there were missing data to be considered.

RAIMUNDO: Is that known by now?

SERGIO: I mean...

RAIMUNDO: Do you remember the extract from the Informe Especial report? What did it say?

SERGIO: That we do not know well.

RAIMUNDO: (...) Perhaps, it is not that they are hiding information from us, it simply is that there are things we do not know.

SERGIO: So, if we do not know those data (...) we are more certain not to build it (the power plant), because at some point we will regard those data (...)

...

CRISTIAN: It is also very likely that the companies having problems in Quintero had ignored problems in data. Is that an option? (...)

Students insist in possible problems of *data*, implying that if we counted with more and better numerical information, a disaster could have been avoided, thus restricted to the notion of risk as in risk assessment, without recognising that decisions are made under uncertainty (indeterminacy, unawareness) as well (Beck, 2000, p. 216). In part, students reducing the unknown to miscalculations or lack of better data can be explained by the *implicit didactic contract* (Brusseau, 2002), by which students respond based on what they believe their expectations are in the context of a mathematics class. More importantly, this contract is established by the way the didactical sequence has been arranged, i.e. by primarily assuming that the probability and risk of an environmental disaster can (and should) be computed.

7.1.2 Computing probabilities

The first part of the sequence consists of revising the company's calculation of the probability of an environmental disaster to happen. In their definition, an environmental disaster would occur if all of the following events take place simultaneously:

A: The plant is working and not on shutdown for inspection by the environmental SEREMI[1].

B: There is at least a declared state of environmental pre-emergency.

C: Poor quality coal is being burned.

D: Emission detectors on chimneys are defective.

The company's calculation gives a probability of:

$$p = 0.000205 \approx 0.02\%.$$

Besides the resulting probability p, the company attached the following information with which it made its calculations.

> Air quality in the Marga Marga province varies with ventilation conditions and, according to it, Valparaíso's environmental SEREMI inspects emission levels. Days in a typical year distribute as reported in Table 7.1.

SEREMI inspects \ Air quality	Good	Regular	Alert	Pre-emergency	Emergency
Yes	10	35	20	1	2
No	92	110	73	14	8

Table 7.1: Air quality for the power plant task

> The power plant would work with coal that comes in different quality levels, which generally distribute as shown in Table 7.2.

Good	Intermediate	Poor
40%	30%	30%

Table 7.2: Coal quality for the power plant task

> Chimneys count with particle material detectors. Revisions and calibrations take place once per week. On average, only eight out of those inspections reveal that detectors are not functioning correctly and hence they are replaced.

Let us call $ED = A \cap B \cap C \cap D$ the event of an environmental disaster. From this information, the company's (erroneous) calculation assumes all four events to be independent, and the detectors' checks to be daily.

$$\begin{aligned} \mathbb{P}(\text{ED}) &= \mathbb{P}(A \cap B \cap C \cap D) \\ &= \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D) \\ &= \frac{297}{365} \cdot \frac{15}{365} \cdot \frac{30}{100} \cdot \frac{8}{365} \\ &= 0.000205 \approx 0.02\% \end{aligned} \tag{7.1}$$

[1] *Secretaría Regional Ministerial* (SEREMI) stands for Ministerial Regional Secretariat.

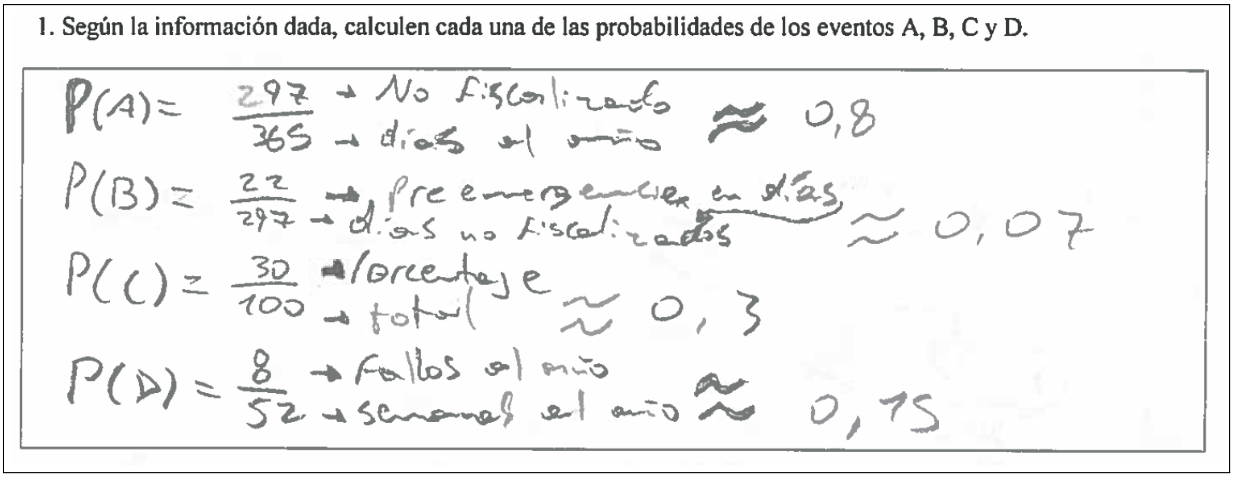


Figure 7.2: Manuel's computation of each probability

This calculation was not shown to the students. Instead, we lead them in the worksheet to compute probabilities separately and to speculate about the calculation and assumptions that lead to $p = 0.000205$. However, only two out the five groups reached to Equation 7.1, the rest either left a blank space or reached a different result by speculating only the error in $\mathbb{P}(D)$. A general takeaway is that, by asking students to make their calculations first, the errors in Equation 7.1 are not easy to imagine, after they have reasoned in a different way directly from available data.

The first instruction in the worksheet we handed out is to compute the probabilities of each one of the events A, B, C and D separately, as can be seen on Manuel's answer on Figure 7.2. The case for this student shows how they rely on the frequentist approach to probability. For events A, B and D, this may seem trivial, since the information is provided precisely in terms of frequencies. However, the information about coal quality is given as a percentage. Still, $\mathbb{P}(C)$ is stated as a fraction between integer numbers. Let us observe each of these probabilities.

7.1.3 Conditional intuition

The calculation of $\mathbb{P}(A)$ is straight forward, and it is the first exercise for this reason; to begin with something simple. Students had no yet studied double-entry tables, but Table 7.1 appears to be self-understood:

RAIMUNDO: You asked me whether you can write on the information sheet or not. Why, what is it that you want to write?

MANUEL: I was summing up.

RAIMUNDO: What did you find by doing that?

MANUEL: (...) These are the days with... [*points to the row labelled as 'No' on Table 7.1*] so I needed that. Since it is in terms of number of days, it should add up to 365.

RAIMUNDO: What added up to 365?

MANUEL: No. That is the total.

Interestingly, Manuel finds no need to verify and count the total number of days in Table 7.1, as his contextual understanding of the information given presumes it.

The calculation of $\mathbb{P}(B)$ should be as simple as $\mathbb{P}(A)$ by reading from the same table, but the trick is to realise that the event B defines "at least" a state of pre-emergency, which is intentionally omitted by the company in their calculation of p. Therefore, the expected computation should be:

$$\mathbb{P}(B) = \frac{\text{Number of days under pre- or emergency in a year}}{\text{Total number of days in a year}} = \frac{1+2+14+8}{365} = \frac{25}{365}.$$

However, Manuel's calculation is different, as he misinterpreted $\mathbb{P}(B)$ as $\mathbb{P}(B|A)$ instead, i.e.

$$\mathbb{P}(B|A) = \frac{\text{Number of working days under pre- or emergency}}{\text{Total of number of days working}} = \frac{14+8}{267} = \frac{22}{267}.$$

From the point of view of the instruction, this calculation is incorrect. However, in the context of the purpose of the calculation as a first step towards finding $\mathbb{P}(A \cap B \cap C \cap D)$, it is correct.

Groups that followed the instruction and computed $\mathbb{P}(A)$ and $\mathbb{P}(B)$ separately tended not to spot the assumption of independence between A and B as problematic and computed their version of $\mathbb{P}(ED)$ accordingly, i.e. by multiplying $\mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D)$.

7.1.4 What is the chance experiment, after all?

The information given on Table 7.2 makes the computation of $\mathbb{P}(C)$ simple, yet its meaning, ambiguous. By revising students worksheets, they simply extract the information as a percentage and interpret it either as a fraction or decimal number, with no further questions (e.g. Figure 7.2). However, close to the end of the joint reflection on Session 3, one student interrupts another group presenting. He seems to be at unease:

CRISTIAN: I wanted to ask a question about the coal thing.

RAIMUNDO: Yes.

CRISTIAN: Eh... That is 100%. 30% of the coal is of poor quality. Eh... That [percentage] of coal is – does it correspond to what is processed in one day, or a determined [period of] time?

RAIMUNDO: That is a very good question...

CRISTIAN: Because if it is 30% – I do not know, I will make a supposition – if this refers, for example, to one workday in the power plant. In one day, it can happen a 30% of – I do not know how...

RAIMUNDO: (...) [*To everyone*] I mean, your classmate's question – if I understood it right – is... For example, 30% coal is of poor quality. Does it mean that 30% of the days only poor-quality coal is used, or that in one day, 30% of the coal used is poor quality?

CRISTIAN: That is where I was going!

RAIMUNDO: And that changes the calculation, which...

CRISTIAN: If every day a 30% of poor-quality coal is used, then that is multiplied by the number of days in a year and that is...

RAIMUNDO: (...) it means that poor-quality coal is used every day. So that would not be 30%, but 100%, and the probability [of a disaster] would not shrink.

Cristian is locating the ambiguous character of the information given trying to interpret $\mathbb{P}(C)$, then thinking aloud to make it into an inquiring question, which I reformulate to help him make his point. What is relevant and follows from there, is the realisation that it could be the case that poor-quality coal is used every day, then the information does not affect reducing the probability of a disaster, i.e. $\mathbb{P}(C) = 1$.

This episode is coded into the category of the "theoretical character of probability" (Borovcnik, 2017); Cristian's questioning is not about how to compute a probability, but about its meaning. Except for event C, all other probabilities are computed from an FQT perspective, i.e. understanding probability as *a posteriori* relative frequencies of repeated chance events. For example, $\mathbb{P}(D)$ represents the relative frequency of revisions that revealed defective particle detectors. The question that remains unanswered is: what is the repeated "experiment" that 30% of its realisations resulted in the use of poor-quality coal? When discussing this episode with Bernardo, he had an additional take on the controversial 30%, which I had not thought through before. When situating the information in context, it is questionable to assume the quality of coal not to affect the air quality:

> BERNARDO: No, yes, yes. I remembered [the question] because that generated conditionalities. Because if there is 30% of poor quality [coal], the probability of existing a pre-emergency or emergency is larger [*referring to Table 7.1*]. That will produce, obviously, that all calculations done are quite conditioned. And the condition is, firstly, that all that was done is wrong. Because it was worked under the assumption that every day it was burned that way.

In terms of the analysis, Bernardo's input situates the episode above in the category of "theoretical character of independence" (Borovcnik, 2017). Indeed, the information given on both Tables 7.1 and 7.2 are assumed to be independent. No student spotted this, and it may be due to the visible way A and B are associated as shown on Table 7.1. On the other hand, an association between C and events A and B can only be conjectured by the context.

As for event D, none of the groups fell for the trap. They identify quickly that revisions are done once per week and that this feature must be reflected in the number of total cases of their probability calculation. Moreover, in the case of Manuel's group, it is the only error they speculate the company had made, as it is asked on the second question of their worksheet: "How is it that the company IS Pagüer reached the result $p = 0.02\%$? (Figure 7.3). Other groups left this second question on the worksheet unanswered. The third instruction was to compute their own version of probability p, making appropriate corrections. Most students, in turn, did respond to it.

The purpose of Session 1, other than setting the scene for the general inquiry, was to scaffold the computation of p with more structure than in previous chapters. The idea was to compute each of the individual probabilities $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(C)$ and $\mathbb{P}(D)$ and fall into the same "traps" as the company, thus verifying their result on the second question. The third instruction should have been a surprise for them, a first encounter with the possibility that their productions could be different. However, that was not the case, as students did not fall into (all) the traps. Their intuitive understanding of Table 7.1 as a double-entry table had them compute appropriate conditional probabilities without knowing it. Reflecting upon the FQT meaning of probabilities prevented a wrongful calculation of $\mathbb{P}(D) = \frac{8}{52}$ as the relative frequency of an experiment that takes place weekly. Speculation about possible wrong ways to compute $\mathbb{P}(ED)$ was not achieved after their discussion and reasoning. Therefore, most of them skipped the second instruction and jumped to answer the third.

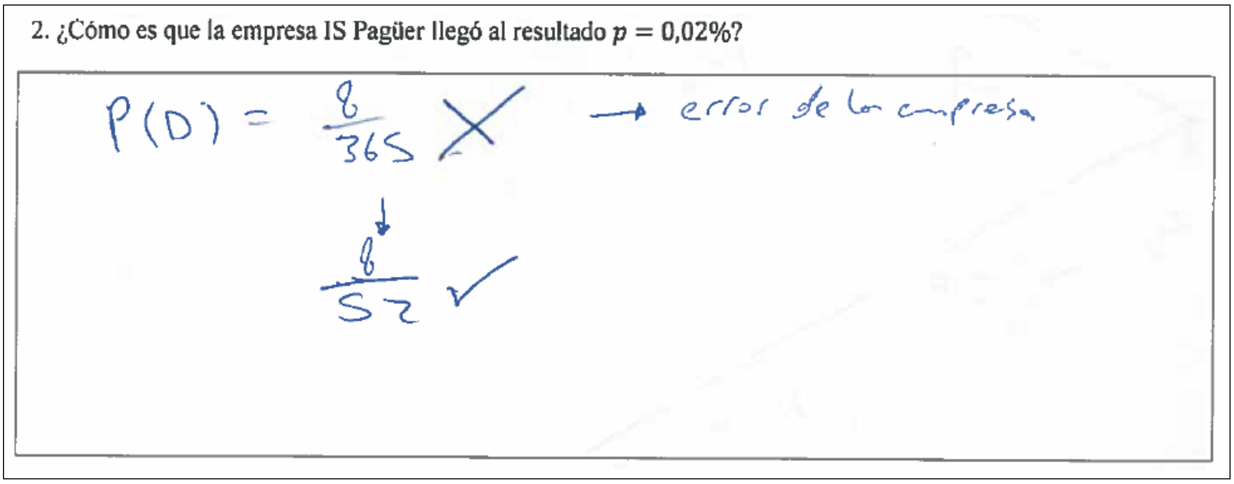


Figure 7.3: Manuel’s statement on the company’s error

By the end of Session 1, the scene has been set in the context of the general inquiry and students have produced their own versions of p, which they shared to the whole class. However, there is no consensus on its result; some of them still fell for the independence trap between A and B. One of the lessons learned from previous chapters is the need to make joint reflections and identifying mathematical issues in the plenum, in order to move forward on the inquiry. That is our plan for Session 2.

7.2 Session 2: The long term

The second session has two purposes: to find a consensual revised calculation of $\mathbb{P}(ED)$ and to extend the probability of an environmental disaster to a determined time span. As teacher and researcher, our intention is for students to compute the probability in such a way it calls for the binomial model. The keys to making that step are, firstly, to understand what does the computed probability means (i.e. each day as a trial with $p = \mathbb{P}(ED)$). From that point, establishing a fixed number of independent trials (e.g. $n = 365$ in one year) would lead to a much larger probability of disaster in a longer-term.

For the first goal of the session, students from all groups share their insights about the calculations. These are crystallised in a new worksheet with four questions:

1. Are all events A, B, C and D independent from each other? If not, which events are conditioned by another?

2. If there are non-independent events, what is the probability of them happening at the same time?

3. Revise the result obtained during the previous session, i.e. the probability of unleashing an environmental disaster.

4. What does that probability mean in the context of the instalment of the thermoelectric power plant? Argue whether the probability is low enough to declare the plant safe.

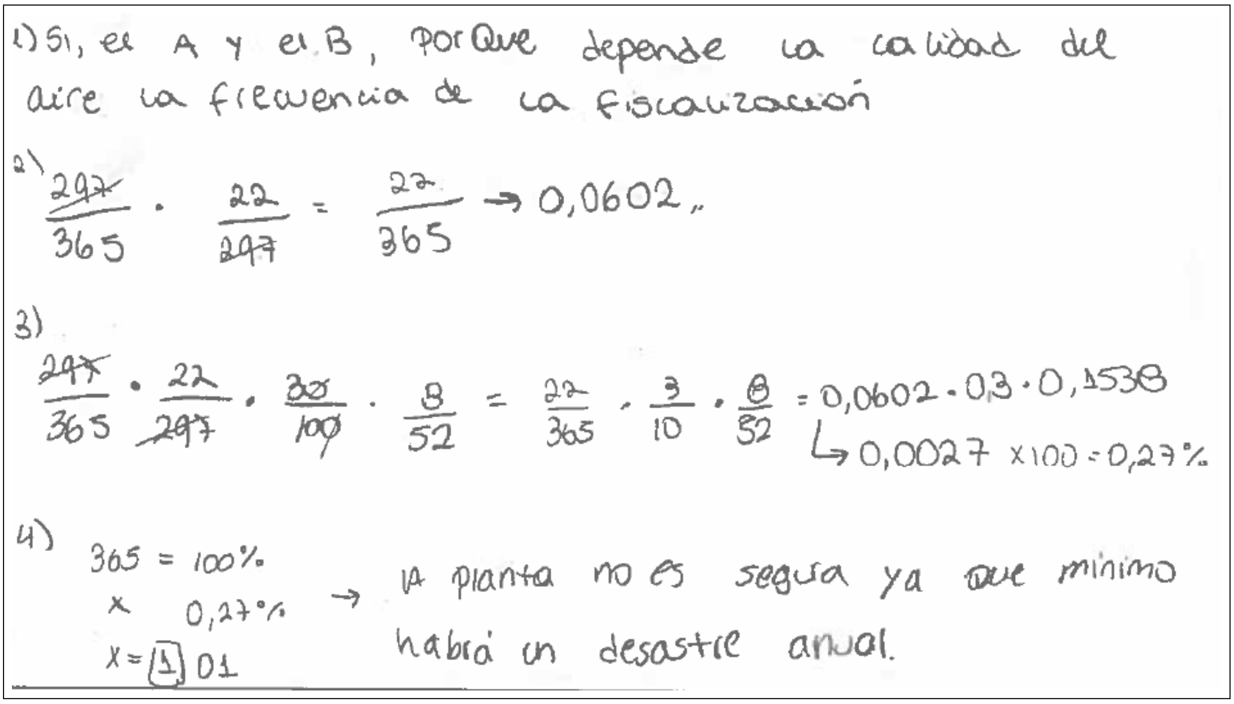


Figure 7.4: Martina's answers to first part of Session 2

Figure 7.4 shows an example of students' answers. In question 1, Martina asserts that "A and B (are not-independent), because the frequency of inspections depends on air quality". The answer to question 2 has the expected numerical results, omitting probability notations, since we have not introduced a formal notation for intersection of events nor conditional probabilities so far. Question 3 is answered as in Equation 7.2 and expressed as a percentage. We can see how a series of questions can help students to crystallise mathematical ideas, but we missed the opportunity to generalise them with proper notations. The students' corrected calculation could be formalised as follows:

$$\begin{aligned}
\mathbb{P}(ED) &= \mathbb{P}(A \cap B \cap C \cap D) \\
&= \mathbb{P}(A \cap B) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D) \\
&= \mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D) \\
&= \frac{267}{365} \cdot \frac{25}{267} \cdot \frac{30}{100} \cdot \frac{8}{52} \\
&= 0.00278 \approx 0.28\%
\end{aligned} \tag{7.2}$$

Question 4 is meant to start triggering students' arguments for a decision. In Martina's case, she is misinterpreting the resulting probability as a proportion. In her reasoning, the probability for an environmental disaster to occur in one day is 0.27% (she rounded 0.00278 down), which proportionally is equivalent to 1.01 days in a year, stating that "the plant is not safe, since there will be a minimum of one yearly disaster". This proportional trap plays a significant role in the unsuccessful attempt to construct a probability for a disaster in the long term, as I pinpoint in the following subsection.

7.2.1 Probability again

During Session 1, probabilities of all events A, B, C and D are interpreted and calculated from an FQT perspective, i.e. as relative frequencies. Is that so that even $\mathbb{P}(C)$ is turned into a fraction from a percentage (see Figure 7.2) and its ambiguity is explored by asking what the repeated chance experiment that leads to a 30% of "success" is (see the dialogue in subsection 7.1.4). However, this view of probability as a ratio has a counterpart: the illusion of linearity.

For the most of it, utterances coded within the theme of the "theoretical character of probability" (Borovcnik, 2017) during the final joint reflection on Session 3 follow a similar pattern to Martina's answer to question 4 (Figure 7.4). To mention a few:

PEDRO: (...) there is a 0.2% chance, which can be seen more or less as 17 hours (...) within a year. That is it.

...

GASPAR: (...) there was a 0.23% chance for it to occur in one year. So we multiplied by 10 to get the percentage (probability) for it to occur at least once within ten years.

...

MANUEL: (...) the probability, in 10 years, for an environmental disaster to occur. That gave us 0.2%, which is approximately 73 days.

RAIMUNDO: How did you do it?

MANUEL: Eh... we made a rule of three. I mean, because there are 3650 days in (...) 10 years. Then, with the (...) 0.2%, it gave us 73 days.

What is happening in those statements is what is known in the literature as the "linearity trap" or "illusion of linearity", i.e. the improper use of proportional reasoning. Employing in-depth interviews, De Bock, Van Dooren, Janssens, and Verschaffel (2002) reveal how this trap is not only due to a firm belief among students on the proportional relation between quantities but also because of ingrained habits influencing basic mathematical conceptions in the course of schooling. Van Dooren, De Bock, Depaepe, Janssens, and Verschaffel (2003) go further and explore the illusion of linearity in probabilistic thinking as a common denominator explaining several misconceptions, in the broad scope from great mathematicians in history to school students. Moreover, their study implies that these mistakes are pervasive despite formal instruction in probability.

In discussion with students, I attempted to make them see the flaw in this type of reasoning, by challenging them into the case of a probability larger than 1. The illusion of linearity is so pervasive, that, in return, it challenges the interpretation of a probability:

RAIMUNDO: (...) But, there was an interrogation I posed to you.

GASPAR: You asked what would happen if we considered a long time, I think 100 or 1,000 years, which leads us to 270% (chance), which is more than 100%.

RAIMUNDO: OK, then, how would we interpret a probability of 270% in 1000 years?

GASPAR: We say that 100% (chance) means it (environmental disaster) happens once, 200[%] means twice. So it would be 2.7 times.

As much as the literature suggests that this heuristic is deep-rooted in our habits and beliefs (Kahneman et al., 1982), the tasks can also play a role. For example, the study of Van Dooren, De Bock, Janssens, and Verschaffel (2007) not only shows how the illusion of linearity can be due, in part, to a scholastic effect perpetuated by the exposure to traditional tasks. They suggest that it could be reversed by subjecting students to authentic performance tasks. An awareness of the linearity trap should not be an excuse, but an opportunity to provoke learning in an inquiry-based environment. For once, Bernardo has a simple suggestion:

BERNARDO: (...) they did not involve the concept of a secure event (i.e. with probability 1), which we had been studying already. Then, you cannot have a *super secure* event, no. (...) So the accommodation they did was simply to convince themselves and move on with the exercise.

We did not take the time to build up from this conflict, and challenging them to question a probability result larger than one was not enough. There is a practical consideration for this pedagogical decision. When starting our Session 2, we realised the classroom was to be used for another activity. Solving the problem with the school, changing to another room and setting up the groups took us around half an hour.

We could have used that opportunity to open the discussion and formalise a probability that follows the binomial model, just as we collected ideas and found a consensus to find $\mathbb{P}(ED)$. From Bernardo's teaching plan, the binomial distribution was not supposed to be covered, so we relaxed the eagerness to formalise it ourselves and expected students to provide their own insights. Perhaps, too relaxed. Assuming that the daily probability of an environmental disaster $\mathbb{P}(ED) = 0.00278$ (see Equation 7.2) is the same for every day and that all days' occurrences of ED are independent of each other, the probability of at least one environmental disaster to occur within one year (365 days) is:

$$1 - (1 - 0.00278)^{365} \approx 0.638, \tag{7.3}$$

whereas the same calculation for ten years (3650 days) would yield:

$$1 - (1 - 0.00278)^{3650} \approx 0.99994. \tag{7.4}$$

Both calculations, in principle, do not require an understanding of the binomial distribution. It would suffice to apply the probability of the complement and the multiplication rule for independent events properly. According to Bernardo's plan, these notions were already studied and exercised. We did not activate them.

7.2.2 We work with what we have

As already mentioned, we intended to get students to extend the probability of a disaster to occur on a random day to a longer time span, e.g. one or ten years. Despite not being a successful feature in the sequence, some comments during the joint reflection illustrate an opportunity to discuss assumptions behind calculations as in Equation 7.3 and Equation 7.4.

RAIMUNDO: (...) For all calculations – or attempts – we assumed the probability to be the same for every day. But you said that is not necessarily the case, why?

SERGIO: Because of the machinery's deterioration. As machines wear down, they become less efficient and (...) the probability of a disaster increases.

...

MANUEL: (...) Besides, the probability does not consider that the province will always grow: its population, the number of vehicles in circulation, etc., and that affects the probability of an environmental disaster.

RAIMUNDO: How so? Elaborate on it, for I am interested. You tell me the population is growing...

MANUEL: The population grows and thus it produces more pollution. There are more cars, more waste, and that affects too. That would be it.

RAIMUNDO: Yes. That last point is interesting, because (...) we had not considered it before, and it has to do with the fact that each year, each day or each chance experiment, they should be independent of each other and with the same probability of failure or disaster. But it is not [necessarily] like that. They [*to Sergio's group*] mentioned the machinery and here [*to Manuel's group*] they said the community itself is changing. Therefore, (...) those air quality indexes we had on the table (7.1) are probably not the same every year. They are probably changing.

Both Sergio and Manuel are advocating against the computation of the probability in the long term, not from the mathematical entry points of "are we doing the calculations right?" or "are we doing the right calculations?" (Skovsmose, 1992, 1994), but from the perspective of the appropriateness of the application to the purpose in mind. Whichever the approach is taken, the assumption of an equal likelihood for the environmental disaster to occur every day is doubtful.

During the interview, Bernardo approaches this episode imagining more assumptions that could be hunted and questioned:

BERNARDO: (...) and someone could have said: "Hey, but during wintertime, the air quality and wind conditions are different", all those sorts of things. So, with that question, which no one else took over, could have provoked a huge avalanche. Because it could have invalidated the whole assumption behind the exercise.

Bernardo is criticising the task itself. As it is formulated, the only way to make risk computable through probabilities is to make assumptions that do not hold when put into context. This issue can be seen as a further potential of the task for students to reflect. However, a further tension emerges if our goal is to advance with the inquiry or reformulate it in future opportunities:

BERNARDO: Because, first, we begin to add more variables and collect more things. Immediately, what the company presented to us falls apart. And then, we have to produce a new data collection, compute new probabilities, and we did not have that. And if we tell them "let us work with what we have" (...) it would produce too much ambiguity, because they are not independent data.

The tension is situated between the *inquiry approach* and *pragmatic* considerations. On the one hand, students could do independent work to collect their own data. This search is, for example, what students do in a task to investigate the safety of nuclear power plants as reported in Radakovic (2015b). It allowed them to not only estimate empirical probabilities but to evaluate the credibility of sources of data as well. On the other hand, Bernardo acknowledges that producing a new data collection is not possible due to time restrictions and the availability of computers, for which we did not plan. Therefore, we "work with what we have". Finally, we come back to the *authenticity* feature of the inquiry approach. Working with what we have, based on knowingly flawed assumptions, seems to miss the point of investigating an issue anchored to a real event.

7.3 Session 3: Decision making in the risk society

In Section 3.1 I argued for how the development of probability and statistics is intertwined with the practice of decision making under risk or uncertainty. For Borovcnik (2015), risk and decision-making define the "logic of probability", and hence probabilistic models and applications play a relevant role in the risk society. Beck identifies as one of the most relevant sociological issues "the socially very relevant distinction between risk *decision-makers* and those who have to deal with the consequences of the decisions of *others*" (Beck, 2000, p. 213, emphasis in original). As a crucial part of the general *problématique*, we intend in our didactical sequence to join ideals of democracy in the sense of deliberative mathematics education (Valero, 1999). We fantasise about short-circuiting the tension pointed out by Beck, by having students, who are locals, to be acting and reflecting as decision-makers.

The third and last session is meant for students to finish and clarify any pendant calculations and to rethink the general inquiry about the instalment of the thermoelectric coal power plant. After summarising some of my and Bernardo's insights from students' work on the previous sessions, we ask each group to prepare a presentation in the plenum, using the following three questions as guidance:

1. What is the probability of unleashing an environmental disaster?
2. What is the probability of unleashing an environmental disaster within ten years?
3. Should we install the thermoelectric power plant or not? Consider all relevant factors.

We are explicit about asking students to be as precise as possible in the first two questions, in terms of interpretations, to address their ambiguity. Is the computed probability during Session 1 the probability of precisely one disaster or at least one disaster to occur? Is it during a single day or in the course of a full year? In the third question, they should bring all considerations to the table, other than probability calculations. The complexity and subjectivity of decision making in socially relevant issues is at the core of the intended reflections.

One of the categories coded in the analysis is "the ability to separate reflecting on a problem and making a decision", from Borovcnik's (2017) notion of probabilistic thinking. Students unanimously reject the instalment of the power plant in their province, despite recognising that the probability for an environmental disaster to be unleashed (even in the long term) is very low. What becomes more interesting is the variety of arguments against the power plant which are separated from computing a probability, and illustrate some of the sociological concepts of the risk society.

7.3.1 Risk is more than probability

As seen in Section 3.1, one of the major transitions towards a mathematical notion of probability was that from conceding the role of decision-maker to chance as a manifestation of God, towards acknowledging the role of chance when one is the autonomous decision-maker. Early driving problems were mostly of pricing of gambles, whereby not only the likelihood of different scenarios was at stake but also their impact, which later was rethought as expectation and utility. Let us call $ED = A \cap B \cap C \cap D$ the event of an environmental disaster. In a rather simplistic view, the risk $R(ED)$ can be expressed as $C(ED) \cdot \mathbb{P}(ED)$ (risk = accident × probability, writes Beck (2000, p. 213)) of an event ED is the product between its probability $\mathbb{P}(ED)$ and its consequence $C(ED)$, whether it is a cost or a benefit. Alrø and Skovsmose (2002, p. 228) argue that "if $\mathbb{P}(ED)$ is terribly small whereas $C(ED)$ is big, could then the mathematical risk calculated become misleading? Could it be that even though $R(ED)$ appears small, the risk is not acceptable nevertheless?"

The problem can be approached by both questioning the meaning of $\mathbb{P}(ED)$ and whether $C(ED)$ is well defined so that the product can be well defined. $\mathbb{P}(ED)$ may be interpreted from an FQT perspective, considering every day or year as a chance experiment. However, the decision is not repeated, and "one-off decisions and repeated situations do require a different logic, which confuses and might provoke a focus on subsidiary information" (Borovcnik, 2015, p. 137). We cannot rethink the instalment of the plant every day. Concerning the consequence, one of the provoking statements in the task is the fact that the company refers only to a probability and does not speak of risk. Therefore, $C(ED)$ is subject to subjective perceptions and, more importantly, "the difficulty with such societal risks is that small probabilities are combined with threats of an enormous damage" (Borovcnik, 2017, p. 1503). The consequence or impact is mentioned as an environmental disaster, but it is not quantified, and it may, in fact, not be possible.

From the analysis, some of these issues of small probabilities are part of the joint reflection, coded accordingly. A query on the word "impact" during the interview illustrates Bernardo's impressions about students' reflections.

SERGIO: (...) despite the result we obtained, should this (power plant) be convenient? We think not, due to the fact that even if the probability is low, the benefit is also little.

Sergio's group takes a cost/benefit analysis approach. Moreover, their presentation considered the possibility of reducing the cost of electricity by allowing the plant to be constructed, only to question it themselves, since:

SERGIO: (...) energy is accumulated and is later distributed [throughout the whole country] (...) for which it is not convenient to destroy the zone.

Bernardo remembers well how the considerations of cost and benefit in decision making are an opportunity to develop further from probabilities:

BERNARDO: (...) They started to analyse the pros and cons. Within the cons and explanations, they do mention it (impact) (...) pointing to the fact that, regardless of finding a low probability, that impact could be very high.

In other words, a decision implies costs and benefits under different scenarios and merely calculating probabilities is not enough. Moreover, Bernardo emphasises that impacts could eventually be *too* high, which students also locate while thinking aloud:

PEDRO: (...) What we realised is that despite there existing low probabilities – how do I say it – low joint probabilities, they produce a big disaster. So – I do not know – somewhere there is a 0.2% chance (...), but it does exist the case for that (four events) to occur in the same day, it will produce a very... a giant disaster, which is not ideal.

...

PEDRO: We also thought and realised that despite there existing a low probability, ideally there should be no probability for an environmental disaster to occur (...)

Pedro and his group are arguing from the perspective of immeasurable catastrophic damage, wherein only a null probability of disaster is acceptable. After their intervention, I challenged them by asking twice whether it is possible to yield a probability of 0. Both inquiries are followed by dead silence. I reformulate by asking about revising events A, B, C and D, and two suggestions are given: to simply not use poor-quality coal and to check emission detectors regularly.

During the interview, Bernardo recognises that the considerations of impact are not formalised during the task, but it becomes relevant for making the final decision.

BERNARDO: (...) Expectation, of course, is not elicited. However, there is something (...) towards the end, during the third session, when they say "well, should we install the plant or not?", many at least insinuate an understanding that it is a matter which not only depends on the probability for a disaster to occur but what is its impact. And they do not multiply it, because how can you calculate the impact of an environmental disaster?

...

BERNARDO: That is something that is not seen within this content, let us say. For now, it is only about computing probabilities. However, considering a random variable – e.g. impact – that impact could be gigantic. So it somehow insinuates that something else is missing to study this case.

Bernardo's realises a tension between *exemplarity* and *pragmatism*. In Section 3.3 I have pointed out how notions of decision making are invoked in the Chilean 11th-grade curriculum as part of the justification. The power plant instalment decision is exemplary for concerns of the risk society (critical perspective) and the use of probability in decision-making (instrumental). Bernardo's year plan for the class does cover many probabilistic ideas, but notions of random variable or expectation are not included, which he pinpoints to be necessary for this case study. The official curriculum does include the concept of (discrete) random variables, and students are expected to represent Bernoulli, and binomial distribution functions (MINEDUC, 2015b, p. 121), but, because of pragmatic considerations, he did not include them in his plan,

BERNARDO: (...) for the simple reason that they were already behind in curricular coverage.

However, the fact that students evoked notions of impact beyond decisions shows the potential of the task for developing notions of risk composed by both probability and consequence, and to reflect upon whether these consequences can be mathematised.

In the analysis, the duality of impacts that can or cannot be measured explains why the coding of small probabilities is often overlapped with the "ability to balance between psychological and formal elements", as a category of probabilistic thinking (Borovcnik, 2017). Nevertheless, that category goes beyond the assessment of a disastrous scenario.

7.3.2 Factual and value statements

Decision-making in risk scenarios cannot be simply conceived as a rational and consistent endeavour. As seen in Section 3.2, a whole research programme is dedicated to explore and characterise archetypal strategies or shortcuts that do not necessarily match what could be considered a rational deliberate alternative (Batanero et al., 2016), described as primary and secondary intuitions (Fischbein, 1975) or as systems 1 and 2 of mind (Kahneman, 2011). Research has shown how stable these heuristics are, even after formal training (Garfield & Ben-Zvi, 2007). Alternatively, Borovcnik (2017) suggests that, in contexts of risk, a probability literacy should involve balancing competing intuitions, including the duality of psychological and formal elements.

Anchoring

One of these strategies is found going through students' arguments to ground their decisions, namely the anchoring heuristic. Anchoring is a heuristic through which recent information influences a probability judgement (see, e.g. Tversky & Kahneman, 1974). For example, a recent plane crash on the news would make us perceive the (subjective) probability of plane crashes to be much higher than before we became aware of the crash. Coded as "psychological vs formal" in the analysis, this is a typical case of anchoring in the joint reflection:

SANDRA: It is just that, if the probability of it happening in Quintero and Puchuncaví is the same for it to happen here, why is there declared a zone of environmental saturation? I mean, the fact that it (the probability) is small, it does not mean that nothing will ever occur here.

To give some context, this comment by Sandra was given right after her groupmate Sergio had spoken in terms of a cost/benefit analysis. It can be considered as a formal aspect, regardless of the assumption that those costs and benefits can be measured in reality. Sandra is questioning the meaning of a small probability, anchored in the recent happenings in Quintero and Puchuncaví, i.e. her subjective judgement of probability is influenced by recent news. Moreover, the case of environmental saturation was at the core of setting the scene at the beginning of the sequence, and one might say that inducing to bias is bad practice. In my view, this is an opportunity for students to confront their perceptions of their calculations.

Values

Estimating the impact or likelihood of a disaster is not the only conflict between intuitions. One of the characteristics to acknowledge in the risk society is that risk statements are both factual and of value. Even if we assume calculations of probability and impact to be reliable, risk perceptions are subjected to what one considers to be tolerable or not. Therefore, "in a risk society, the question we must ask ourselves is: how do we want to live?" (Beck, 2000, p. 215).

That is the case, for example, of the chapter on risk management and decision-making about sustainable development in the Intergovernmental Panel on Climate Change report about climate change (IPCC, 2019). On the one hand, experts provide their best predictions of consequences of different courses of action, and a level of confidence about these claims. On the other hand, they acknowledge that they are not the ones making decisions. The world will not end, but governments and the general public ought to ask themselves whether we want to live in such scenarios.

This idea appears to be more or less spontaneous among some students during the joint reflection, as their rejection to install the power plant does not only refer to the worst-case scenario:

GASPAR: (...) And it would disturb the place's flora and fauna. Besides, pollution will reach here too, and many people living in this zone would have to leave. Moreover, the older adults – of which there are many – would get sick quicker and life expectancy would decline.

...

RAIMUNDO: Anything else you want to add?

NICOLAS: (...) It does not have to... the word *disaster* is not necessary for the air to be polluted and the environment to be bad. Instead, for example, in Santiago, they have to impose [vehicular] restrictions sometimes. I mean, in Santiago, there is pollution every day.

RAIMUNDO: Right.

NICOLAS: Sometimes it is more, sometimes less. The point is that it is always polluted.

Santiago, Chile's capital city, has a long, never-ending history of air pollution and smog that has been normalised and palliated with sparse restrictions. Nicolás is saying that regardless of the occurrence of a disaster, his quality of life is at stake. He is questioning an underlying assumption in the task. That is, the criterion for deciding on the power plant is reduced to the possibility of a disaster.

Past, present and future

The question of "how do we want to live?" takes part during the interview. As soon as we remember the case that contextualised and set the scene for the sequence, Bernardo points to yet another issue of the risk society, namely how "a threatening future, still contrary to fact, becomes a parameter of influence for current action" (Beck, 2000, p. 221). He was telling me that the zone of Quintero and Puchuncaví used to be an attractive coastal touristic destination.

RAIMUNDO: I have heard – I do not know if as a euphemism – about a sacrifice zone.

BERNARDO: Yes. It (the zone of Quintero and Puchuncaví) is declared a sacrifice zone.

RAIMUNDO: (...) One is either not allowed to live there or bound to deal with a poor quality of life, but we accept it because we have to put industries somewhere.

BERNARDO: The thing about a sacrifice zone... The idea is to declare it *before* people already live there. First, there was the city of Quintero, the people of Quintero, and then all those industries were installed.

Here I encounter a contrast between Beck's assertion and Bernardo's take. For Beck, in the risk society, "the past loses its power to determine the present" (Beck, 2000, p. 213), whereas Bernardo's comments suggest otherwise. He claims that decisions were not made with a threatening future in mind, but the other way around; a province is being declared sacrifice zone *after* an impact has been experienced. However, this divergence can be seen as an opportunity. Beck, Giddens, and Lash (1994) argue that establishing and delineating a problematic situation does not mean one cannot respond to it as well. After all "the theory of risk society develops an image that makes the circumstances of modernity contingent, ambivalent and (involuntarily) susceptible to political rearrangement" (Beck, 2000, p. 221). In that sense, a teaching experience that positions students as decision-makers can be exemplary for a society that reflects upon itself (Beck et al., 1994). Students can take part in decision making upon a virtually threatening future and not only in the form of damage control. That is the imagined purpose of the didactical sequence.

Bernardo builds on beyond the case study. Though he fears to be going on a tangent, I find his statement most appropriate:

BERNARDO: (. . .) We are a reactive country. It means that one could be building something and say “OK, let us do this”. A couple of neighbours oppose, but the rest of the city does not. When do they oppose? When they realise that their harmony has been broken (. . .)

BERNARDO: So we are a reactive country. A country used to, in this case, hard data from descriptive statistics. But those are there once things have already happened. Nothing in this country is made pro-something but always post. Healthcare policy, education...

He holds an opinion in a sociological breadth. For him, as a nation, our normality is making decisions and attempting to solve problems after negative impacts have been experienced. Coming back to the general *problématique*, this is yet another contingent reason in line with the critical justification; we need an education towards precautionary, premeditated, socially relevant decision making. More specifically, Bernardo highlights the limitations of descriptive statistics, giving further grounds to probabilistic thinking in the context of risk to be a focus-point in the classroom, supporting the claim that “the teaching of probabilistic thinking is important and justified in its own right and not simply as a tool to pave the way to inferential methods of statistics” (Batanero et al., 2016, p. 25). The inclusion of probability is then justified for its intrinsic connection to risk and decision making.

The power plant at the centre of the discussion in the learning environment is fictitious, but based on a real project; a gas-fuelled thermoelectric power plant to be installed in the Marga Marga Province, home to Bernardo and his students. The case of Quintero and Puchuncaví rose concerns and rejection on behalf of local organisations, afraid of becoming a new unwanted sacrifice zone. However, the company’s environmental impact evaluation has been approved, and the plant is on its way to be constructed. During the writing of this chapter, the area of Quintero and Puchuncaví reached yet another peak of atmospheric sulphur dioxide (Jara, 2020). In response, and according to a new decontamination plan, regional authorities have activated protocols to lower industrial activity. To be precise, the peak is occurring in April, in the Southern hemisphere’s autumn and start of the influenza season and other respiratory viruses. To be more precise, it is April 2020, amid a severe acute respiratory syndrome coronavirus (SARS-CoV-2) pandemic.

7.4 Back to the research questions

Before moving onto the next classroom experience, I would like to revise some lessons from the “Should we install a thermoelectric power plant?” learning environment, in light of the research questions.

7.4.1 Reflective knowing

Regarding **RQ4** and **RQ5**, is there evidence for students engaging in reflective knowing (Skovsmose, 1994), anchored in topics and abilities of probabilistic literacy in the context of risk (Borovcnik, 2017)? Let me revisit the utterances highlighted in the analysis in light of mathematical, technological and reflective knowing.

Mathematical-oriented reflections

Mathematical knowing is evident in the form of "have we done the right calculations?" in computations of basic probabilities. For example, when Manuel is computing $\mathbb{P}(A)$ supported by Table 7.1, he checks whether he is going in the right direction by summing up a row and dividing by the total of 365 days in a year. Moreover, students also judge the calculation of the company, as seen on Figure 7.3.

Another key example to consider is when students fall into the proportional trap and extend the probability of a single-day disaster:

GASPAR: What would happen if we considered a long time, I think 100 or 1000 years, which leads us to 270%. (chance), which is more than 100%.

RAIMUNDO: OK, then, how would we interpret a probability of 270% in 1000 years?

GASPAR: We say that 100% (chance) means it (environmental disaster) happens once, 200[%] means twice. So it would be 2.7 times.

These examples of mathematical knowing are embedded in the topic of the theoretical character of probability from an FQT perspective. If students also recalled that a probability cannot be larger than 1, then they could have changed course and question their proportional algorithm for the intended purpose.

Model-oriented reflections

During Session 3, most reflections can be considered to be of the technological kind, as they relate to the purpose of computing probabilities in the context, beyond the right-wrong divide. Even when the algorithm of applying proportional reasoning is incorrect, students are still questioning some of the assumptions which would eventually lead to a correct result such as in Equations 7.3 and 7.4. Their reflections about the wearing of machinery and population growth put in question the assumption of holding the same probability on each day or year.

Another relevant example is Cristian wondering: "That is 100%. A 30% of the coal is of poor quality. Eh... That [percentage] of coal is – does it correspond to what is processed in one day, or a determined [period of] time?" Computing $\mathbb{P}(C)$ is not an issue for mathematical knowing, since its computation is given. However, its function in understanding the risk of an environmental disaster is ambiguous. Bernardo reinforces this impression by claiming that "all that was done was wrong", i.e. that the calculations may not help us solve the problem.

Technological reflections also imply asking whether mathematics is needed at all. When deciding against the instalment of the power plant, Nicolás questions the way the inquiry is posed:

NICOLAS: (...) It does not have to... the word *disaster* is not necessary for the air to be polluted and the environment to be bad. Instead, for example, in Santiago, they have to impose [vehicular] restrictions sometimes. I mean, in Santiago, there is pollution every day.

His reflection goes certainly beyond mathematical algorithms and their appropriateness for solving the task. He is questioning the fact that assessing the risk of a disaster is relevant at all for decision making.

Technological knowing is also evident by the fact that a low probability seems unsatisfying for students. Sergio claims that "even if the probability is low, the benefit is also little", while Pedro says that "ideally, there should be no probability for an environmental disaster to occur". These reflections question the appropriateness of merely computing a probability for deciding since considerations of impact are missing.

Context-oriented reflections

Reflective knowing implies questioning the formatting power of mathematics. From my point of view, this type of knowing is not evident from the analysis, but its potential is illustrated by connecting emerging reflections to the issues of decision making in the risk society.

I attempt to prompt the risk-uncertainty problem, i.e. what can and cannot be computed, by making connections to the environmental disaster in Quintero and Puchuncaví and its causes, namely:

SERGIO: That we do not know well. (...) So, if we do not know those data (...) we are more certain not to build it (the power plant) because at some point we will regard those data (...)

CRISTIAN: It is also very likely that the companies having problems in Quintero had ignored problems in data. Is that an option?

I could not claim that they meant it, but Sergio and Cristian reflect along the line of realising that some risks are not possible to compute, and risk assessment may give a sense of false security. They imply that given that other disasters did happen by causes that were not calculated, making calculations does not guarantee any safety for a new power plant. It becomes even more evident when Sandra says that "the fact that it (the probability) is small, it does not mean that nothing will ever occur here".

When justifying their decisions, students raise many factors that go beyond risk assessment. In the analysis, this is coded as the ability to balance between psychological and formal elements (Borovcnik, 2017) and relate to how risk statements are both factual and of-value (Beck, 2000). For example, Gaspar brings issues of local flora and fauna, emigration from the zone and older adults' health. This mention can be easily regarded as auxiliary information people use to confirm their established beliefs in a risk situation (Pratt et al., 2011; Radakovic, 2015a). However, considerations outside of what can be measured are necessary to reflect upon how mathematical tools limit our perception.

From the analysis, the potential of engaging in reflective knowing by this learning environment is embedded in the probabilistic thinking of separating the reflection and decision upon a problem (Borovcnik, 2017). Another common element to the critical potential lies in how students analyse the fictional power plant of the task with the real case of Quintero and Puchuncaví. Having more than one example at sight may contribute to find commonalities and make them exemplary of a broader phenomenon. This anchoring can help revise the principle of *exemplarity* in future designs.

Bernardo's input in the analysis

The framework I have used for the analysis, developed by Borovcnik (2017), pinpoints themes and abilities related to probability literacy in the context of risk. It allows me to have a broad picture of the emerging topics amongst students' reflections. In order to address the critical aspects related to risk, I have drawn upon Ulrich Beck's conceptualisation of the risk society (Beck, 2000) to connect discussions in the classroom at a societal level. This double-layered framing seems appropriate to address the general *problématique* of coherence, by spotting critical issues that are explicitly connected to high school probability and risk. It raises the opportunity to develop a combined framework for what it means to be both probabilistic and critically literate. In the next chapter, I attempt to do so through the critical statistical literacy framework developed by Weiland (2017).

In Chapters 5 and 6, I discuss the potential of the learning environments to provoke critical reflections in a personal speculative manner, anchored in what *did* happen during the classroom experiences and what, in my view, possibilities to what *could* be the case (Blomhøj, 2006). In this chapter, the interview with Bernardo allows me to revisit the experience and some particular episodes. It also lets us project the potential of the learning environment into the future. A case in point is the problem with the 30% poor-quality coal usage. In my initial interpretation, the point risen by Cristian was a concern about the theoretical character of probabilities, whereas, from Bernardo's input, the issue has the potential of discussing the meaning of conditional probabilities and of eventually questioning the entire validity of the task. Bernardo also points to how, in his view, Chilean society tends to be reactive instead of proactive. Anchored in the declaration of Quintero and Puchuncaví as a sacrifice zone in the aftermath of a series of environmental disasters, Bernardo sees the potential of learning environments guided by a decision-making inquiry to raise awareness about the question of "how do we want to live?" (Beck, 2000) or the separation between reflecting and deciding upon a situation (Borovcnik, 2017), and not to appeal to just do damage control. The interview will also be of significant value in Chapter 8.

7.4.2 As for the design principles

The principle of *pragmatism* should be approached as a collaborative endeavour of *practical organisation* between researcher and teacher (Skovsmose & Borba, 2004). After revising the coding of the interview with Bernardo, I realised that the few utterances in the *collaboration* theme (see Section 4.4) are mostly my prompts that he did not follow in that direction. He did not comment on how collaborative our work was. At most, he remembers how we decided to schedule the intervention according to his plan. This revision supports my impression that my role as a researcher overshadowed his role as a teacher. To be specific, I was mostly interested in letting the students' devolutions inform my research questions whichever the form they take, missing opportunities to clarify and crystallise mathematical ideas. For example, students do not reach a proper calculation of the probability of a disaster in the long term, falling into the proportional trap to go about the inquiry (De Bock et al., 2002; Van Dooren et al., 2003) without our feedback. The desired development as in Equations 7.3 and 7.4 is possible under the assumption of independence and equal probability between disasters over time, but it is not made available to the students. On the other hand, they do reflect on issues that can be framed

– implicitly – as questioning these assumptions, namely the possibility of machines wearing out and population growth. This episode does advance my research agenda, as it illustrates the potential of the learning environment to provoke critical reflections, but it does not contribute to Bernardo’s role as a mathematics teacher.

In hindsight, I interpret this tension as a misunderstanding of the *inquiry approach* as if it were unguided discovery learning. Kirschner, Sweller, and Clark (2006) claim that minimal guidance can only work when learners have enough knowledge to provide their own guidance, which is not the case in this experience. However, Hmelo-Silver et al. (2007) criticise Kirchner and colleagues’ inclusion of problem- and inquiry-based instruction in the set of unguided approaches, since these do include strong scaffolding. On that line, better planning of the learning environment would have considered making more pauses when needed, as we did after the first session to compute $\mathbb{P}(ED)$. Further input from Bernardo’s teaching plan would have given us strategies to crystallise mathematical ideas to advance in the inquiry. Moving from the exercise paradigm to landscapes of investigation implies getting out of a comfort zone and running risks Alrø and Skovsmose (2002); Skovsmose (2011) and, as it is the theme of this chapter, risks entail facing eventual consequences.

Authenticity is another aspect of the *inquiry approach*. In his study concerning the risk of nuclear plants, (Radakovic, 2015a, p. 326) proposes, among future directions of research, to “frame a socially authentic inquiry-based learning instruction placed in the students’ own community”. His understanding of authenticity lies in the distinction between cultural and personal authenticity (Murphy et al., 2006), which I understand as *critical and subjective exemplarities*. Nonetheless, I have responded to this call by situating the inquiry in an authentic event taking place in the students’ province. The counterpart is that data and solution strategies are fictional, missing the opportunity to assess the legitimacy of sources of data in the real case. From the perspective of my research focus, there is no harm, as students engage in discussions as if data were authentic. For example, when Cristian comments that “it is also very likely that the companies having problems in Quintero had ignored problems in data”, it takes place nearly at the end of the experience. I interpret his comment as projecting the “problems in data” we explored with fictional information into the past real case of Quintero and Puchuncaví. The use of non-authentic data did not seem to restrain his possibility to reflect upon an authentic situation. We used fictitious data because, to my knowledge, no technical information is available to compute the probability of an environmental disaster as we intended. This obscurity is another potential for critical reflections, as I will discuss in the next chapter, where, upon the lack of transparent information, students construct their own data, based on authentic sources.

Chapter 8

"How many people attended the students' march?"

> The government yet again belittles the figures of attendees to these type of mobilisations. The government wrongfully measures things it finds inconvenient.
>
> Daniela Manuschevich, geographer, after women's march, 2020.

One exemplary case of mathematics in action is the use of estimates of attendees to demonstrations in order to validate – or invalidate – their cause. Organisers and authorities often provide different estimates reported in the media based on non-transparent methods and frame the results according to their respective agendas. This phenomenon provides an opportunity for students to engage in critical reflections about the use of mathematics in society and experience by their own means the process of elaborating an estimate from the construction of data to their conclusions.

The core of the case is a statistical estimation activity inspired by the *Counting People* task developed in the context of teacher professional development by Triantafillou et al. (2018). During the process of planning, Carla, the mathematics teacher, pointed out that her students had attended the demonstration the learning environment is contextualised in, giving an additional point of interest to the role of context in statistics (G. W. Cobb & Moore, 1997; Moore & Cobb, 2000). Moreover, the task can be supported by formal inferential elements in the 12th-grade mathematics curriculum, whose statistics syllabus includes the estimation of population means and construction of (normal) confidence intervals (MINEDUC, 2015a, p. 115).

The task consists of estimating the number of people participating in a march, and it is made into a didactical sequence by dividing the activities into three sessions:

1. Setting the scene with news articles that show estimates for the number of attendees to the event. Estimating the number of people using a closeup picture (see Figure 8.2).

2. Extrapolating previous results into a broader picture of the entire demonstration (see Figure 8.3).

3. Presenting and discussing results contrasted to estimations from the scene-setting of the inquiry.

In reality, the sequence was carried out in three sessions, of 90, 90 and 45 minutes, respectively, wherein the last ten minutes of the third session were dedicated to the presentation of results in the plenum and final reflections. As mentioned in subsection 2.3.3, the analysis for this chapter is based on a framework for critical statistical literacy by Weiland (2017), and transcription and coding is made onto the last part of the plenum and joint reflection. Using this framework, I attempt to analyse the interactions in the classroom during the joint reflection as evidence for both critical and statistical literacies consolidated together. An overview of the coding is shown on Figure 8.1 as a screenshot from the NVivo project. There are three main outtakes worth highlighting from this overview.

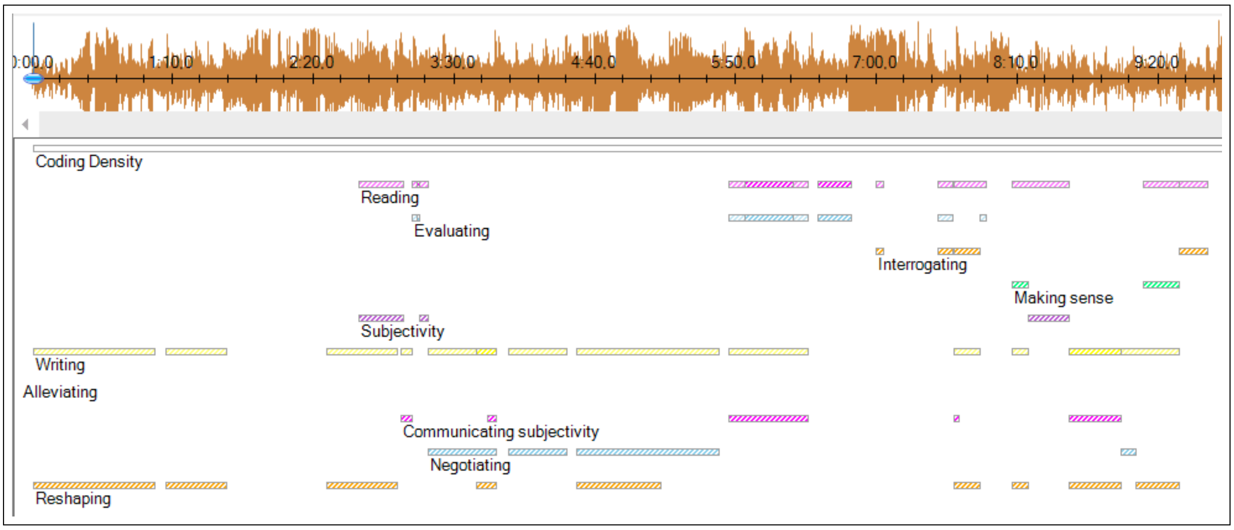


Figure 8.1: Coding of 10 minutes of joint reflection according to critical statistical literacy

First, most reflections pertain to the *writing the world* category of aspects (52% coverage), compared to *reading the world* (24% coverage), illustrating the role that the investigative nature of the task plays in engaging reflections. This share is consistent with the philosophy of Dewey (1938/2015), portraying reflective inquiry as a bridge between learning and doing. Students tend to base their reflections more on their productions than speculating about others' (Alrø & Skovsmose, 2002; Skovsmose, 2011).

Second, within the *writing the world* category, a large extent is coded in the subcategory of *reshaping* (65% of *writing*), the nickname given to the ability to "use statistical investigations to communicate statistical information and arguments in an effort to destabilise and reshape structures of injustice for a more just society" (Weiland, 2017, p. 41). At the same time, no utterances were coded as *alleviating*, defined as "using statistical investigations to alleviate and resolve sociopolitical issues of injustice" (Weiland, 2017, p. 41). This absence can be easily explained by the fact that students refer back to the steps in their statistical investigation as a manifestation of their agency. However, its outcomes do not go beyond the classroom. They may be able to use their work as a way of rethinking aspects of their society. After all, opportunities for alleviating or resolving injustices are not offered during the intervention.

Third, not only *writing* is more prominent than *reading*, but the coding of these overarching categories is not exempt from overlaps. After taking a closer look through the chapter, the evidence suggests that many instances of reading the world are enabled by writing the world, i.e. becoming critical *consumers* of statistical messages (Gal, 2002) is facilitated by playing the role of *producers* (Gutstein, 2006; Weiland, 2017).

For the sake of one of the general research questions of the thesis, namely **RQ5**, it can be seen that it is possible to design and implement a statistics teaching practice whereby students engage in critical reflections. Of course, more interesting is to see some of its instances and how these were enabled during the activities. In what follows, I tell the story of this statistical estimation task, the way it triggers critical reflections amongst students and pinpoints some implications for the general *problématique* of the thesis.

8.1 Session 1: Constructions, sampling and agency

The school where this sequence was implemented is of the subsidised type. A private institution controls it, but it is subsidised by the Chilean State, so parents do not need to pay for tuition, as fully private schools do. The amount of subsidy, however, is subject to a set of rules, including the number and attendance of students. Though classes start at 8:00 am, a leeway is given for the official attendance in a day, cutting at 9:20 am, right before the end of the first block. Students know this rule and know they will be counted as present until that time. Therefore, as Carla explained to me in our previous meetings, we should expect students to be late at our first session belonging to the 8:00-9:30 block. However, according to my field notes, we began the activities with a full classroom by 8:07.

The simplest explanation can be found digging into the interview sub-theme of engagement, within the students' work theme (Section 4.4), wherein Carla points out that:

CARLA: They felt important for being part of a study. Moreover, since we had talked about it at a parents meeting, parents felt important as well. Besides, I believe... I remember they signed something (...) So that is what I felt, that they felt important [*laughs*].

As a manifestation of the Hawthorne effect (e.g. Diaper, 1990), students show up earlier than usual, after the expectation of being part of an investigation. Little they knew that they would be the ones doing an investigation as well.

The demonstration setting the context of the learning environment had taken place on April 19th, 2018, at the beginning of the new presidential administration of right-wing coalition led by Sebastián Piñera. The previous administration, led by Michelle Bachelet, had implemented a series of educational reforms against profit in educational institutions, moving towards tuition-free higher education, and ending academic selection in school admissions. Piñera's educational agenda was based on loosening the established restrictions, and so the march was organised as a warning to power. In the first session, students are presented with the general inquiry: "How many people attended the students' march?" Each student gets a four-page worksheet and two aerial photographs from the march (Figures 8.2 and 8.3). The first page sets the scene displaying extracts from two news articles from the same outlet, before and other after the march, respectively:

April 19th, 2018 (morning)
[PHOTOS] Aerial images of the students' march through Santiago's centre

> *Thousands of students, plus other guilds, met to demonstrate against the government's policies in this agenda.*
>
> This Thursday, April 19th, the first student march of the year is taking place, convened by CONFECH[13], together with the Teachers Union and secondary students' organisations, ACES[14] and CONES[15], expecting to mobilise 60 thousand people.
>
> Source: Meganoticias

At some point of the day, the article was moved to the multimedia section of the news outlet's website and included a series of photographs. I selected the ones taken from the closest (Figure 8.2) and the furthest (Figure 8.3) to give as handouts for the activity. In the evening of the same day, the same news outlet updated the article with the aftermath given by the Regional authorities.

> *April 19th, 2018, (evening)*
> **First student march of 2018: Regional Government estimates attendance at 30 thousand people**
>
> *The demonstration had isolated incidents because of the use of the north side of Alameda avenue by attendants.*
>
> This Thursday, April 19th, the first student march of the year took place, convened by CONFECH, together with the Teachers Union and secondary students' organisations, ACES and CONES, which, as the Regional Government's estimated, mobilised 30 thousand people.
>
> The instance was announced after the Constitutional Tribunal decided to allow for-profit entities to partake as controllers of universities. Thus the motto of this demonstration: "No more profit, no more debt, no more sexist education".
>
> Source: Meganoticias

From there, the overall goal of the three sessions is framed to discuss how estimates can be so different and provide ideas to make our own, to decide on each version's credibility. Students are quick to spot that the organisers' estimate was given before the event and thus an expectation, whereas the authorities did it afterwards. However, none of them is transparent in their methods, so the relevance of the inquiry still holds. We begin by focusing on Figure 8.2 during the first session.

Carla is the mathematics teacher to a 12th-grade class of 28 students, whom we set up to work in five groups. Just as in the previous chapter, a few of these students are given names, for most transcripts pertain to the final joint reflection. The groups are configured as follows:

- Andrés, Gastón and three more
- Silvia, Camila and four more
- César and four more
- Daniela and five more
- Aníbal, Roberto, Eduardo, Bastián and two more

Figure 8.2: Closer picture of the march. Source: Meganoticias

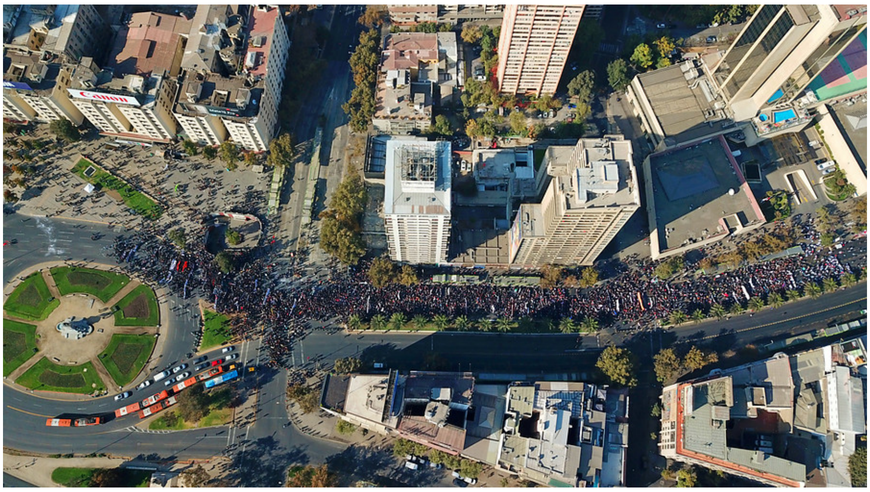


Figure 8.3: Further distance picture of the march. Source: Meganoticias

8.1.1 Data are constructed

Well-known is the phrase “there are lies, damned lies, and statistics”, referring to the manipulation and misuse of statistical information to convey falsehoods as apparently objective facts. This notion is a the core of the popular book *How to lie with statistics* (Huff, 1954), which includes the themes of correlation vs causation, random sampling and distortion of graphs. In his reaction to this book, Joel Best proposes to abandon the illusion of a some perfect “solid rock” statistics to counteract all statistical lies. Instead, he calls for realising that statistics are social constructions (Berger & Luckmann, 1966) he calls “human-made jewels” instead (Best, 2005). The point is not to replace the statistic-or-lie distinction for a calculation-or-construction, but to realise that “every number is inevitably both calculated and constructed” (2005, p. 212). The first part of the sequence shows how constructions begin with data sources, even before collection through sampling, its treatment and reporting. Session 1 is an opportunity to realise that data can be produced (Gal, 2002), by actually constructing it hands-on.

Figure 8.2 is the picture taken from the shortest distance, and so it permits roughly to count heads. With Carla, we guide a plenum discussion to realise how tedious it would be to count everyone on the picture. She leads a dialogue to fish for the strategy we planned to apply, i.e. to use a mesh to divide the picture into frames, considering the perspective it is taken from:

ROBERTO: And we could also see the average *mass* of people, to see how much of that area [*points to Figure 8.2*] they are using.

CARLA: . . . I understand that you want to take an average. But what else could we use, departing from your [*points to* ROBERTO] idea, but without counting mass? [*Points to* DANIELA, *who is raising her hand.*]

DANIELA: Like ROBERTO was saying, we could look at the surface of the street, then look at the square metres each person occupies, and this way look more or less how many are there in the area.

CARLA: OK, I agree, DANIELA. But it turns out that this (the picture) is not in square metres.

ROBERTO: But there we can see a space from here. This [*points to Figure 8.2*] is like a piece of this [*points to Figure 8.3*].

CARLA: What should I do with this picture so we could know how many people are here? Because you [*points to* DANIELA] talked about square metres, but how do I make that square metre appear? Is all of the picture one square metre?

ROBERTO: Drawing small squares, tiny ones... [*general laugh*]

CARLA: [*Smiles*] I find the solution just perfect!

Though several ways of dividing the picture were debated, What was more important to us was reaching a consensus, so we could later collect students' samples and show them empirically how sampling distributions behave according to sample size. This consensus was first to draw a trapeze-shaped perimeter that accounts for the picture's perspective, and divide it into 12×3 frames, as modelled by Carla by projecting Figure 8.2 on the whiteboard and drawing the mesh using a marker as seen on Figure 8.4.

Figure 8.4: Carla's division of Figure 8.2 into 12×3 frames

Students attempt to follow, but as we saw after collecting their worksheets, there was no common agreement reached. On Figure 8.5 three versions are shown. (a) is in hand with the consensus and Carla's model. (b) seems an attempt to follow the model. However, the trapeze is already defined in a different position, and the division is made with ten rows. (c) is a 6×3

division with a wider reach than the others. More surprisingly, these three examples come from the same working group.

(a) (b) (c)

Figure 8.5: Three students’ division of Figure 8.2

Let us remember that this is happening before students begin to take samples and count. Still, we can already observe a source of variability within their constructions of the sample space. This feature would invalidate the aggregation of samples from the whole class later on. A possibility to prevent this issue would have been to hand them out Figure 8.2 with the mesh previously constructed, so everyone is on the same page. However, what did happen helps us to make sense of moving from a statistic-or-lie distinction to a calculation-and-construction awareness (Best, 2005). None of the students is lying and, as we will see later, they all use the same statistical estimation method. Instead, what is happening is that, from the same picture, each student constructs a different version of what is the sample space, represented by the overarching trapeze and the drawn frames as its elements, and the counting will be different form therein. This fact can contribute to the understanding that “numbers do no exist independently of people; understanding numbers requires knowing who counted what, why they bothered counting and how they went about it” (Best, 2005, p. 212).

A hint of the renounce to the statistic-or-lie distinction takes place during the joint reflection on Session 3, in light of the diversity of estimates. The following exchange is prompted by me asking openly about the activity itself:

RAIMUNDO: I wanted to... well, we do not have much time left. We are a bit passed. Still, I wanted to receive a couple of last comments if you want to say anything about the activity itself. How did you like it?

ANIBAL: Cool!

ROBERTO: Good!

ANDRES: It confirmed to me that the more “state” [*does air quotes gesture*] media lie.

RAIMUNDO: OK.

GASTON: That we cannot trust anyone! [*general laugh follows.*] It is just that they take advantage of the situation, I mean, it is like a lie, but . . . [*pauses*].

RAIMUNDO: Disguised?

GASTON: Disguised!

. . .

RAIMUNDO: Is there anything...?

CESAR: More than doubting the media and since it is an own estimate, no one can really tell how many people were at the march, so I would not hold to that suspicion. Since it is an estimate, one can be wrong or not, so I would not take that with me.

Students not only find the activity to be "cool" and "good", but some openly state how it allowed them to reflect beyond calculations. Andrés and Gastón are thinking aloud about a sense of abuse of misinformation hidden in numbers, along the line of lying with statistics, which I reformulate by completing their utterance. César does not conform with this explanation and challenges it by acknowledging that the variation of results may be inherent to the problem and to the fact that estimates are "own", as he experienced it.

In the analysis, the episode is coded as an overlap between two categories. On the one hand, within *reading the world*, I find that, to some extent, they are "making sense of language and statistical symbols systems and critiquing statistical information and data-based arguments encountered in diverse contexts to gain an awareness of the systemic structures at play in society" (Weiland, 2017, p. 41). The potential lies in the fact that he is making sense of what statistical estimates are as constructions and starting to acknowledge the non-neutral role of different actors, including the State and the media. On the other hand, students are "using statistical investigations to communicate statistical information and arguments in an effort to destabilise and reshape structures of injustice for a more just society" (Weiland, 2017, p. 41). Andrés begins by stating that it was their activity the one making him problematise the role of the media in reporting statistical information. The same goes to César, as he emphasises that estimates are one's own. He experienced each of the steps of the task to reach an estimate, so he knows he is not lying, but displaying the results of his construction.

8.1.2 Random sampling as fair sampling

Most of the students' interventions at the beginning make a case for a method that is closer to what urban studies do, by defining some density range measured in persons per square metre. Carla and I had planned for something different, for two reasons. First, units of measurement are relative, so any chosen arbitrary but consistent surface unit would do the job. Second, we originally planned to make use of dice for random sampling in the next step, so students could, for example, divide in 6×6 frames and rolling a die twice two first select the row and then the column of the frame. The plan changed on the go, as we privileged some of the students' ideas and realised that the sampling could be made by generating random numbers with the scientific calculators available in the classroom.

Once the framing of Figure 8.2 is made, the general question is what and how to count. Following the *pragmatic* design principle of curricular adherence, Carla and I intend to use normal confidence intervals as an inferential method. So we lead the discussion fishing for the taking of random samples. I begin by pointing one frame and asking whether it is reasonable to claim that all frames contain the same number of people as the rest. I repeat the question pointing to other individual frames, and the response from students is "no" at unison. The first solution proposed by students is simply to count for every frame and compute the overall average, but it would take longer than the time available to do so. Andrés says we could find a middle ground strategy and Aníbal proposes to take the average between two frames; one of high and another of low density, at least in appearance. So far, we establish the strategy of taking and averaging a *sample*.

Then comes the question of how to take a sample, heading to a discussion around the idea of fairness. There was already a suggestion to cherry-pick two of extreme densities, but we challenge the idea by asking the student whether it is valid to assume that half of the frames would be of high and low density each. That is when Carla comes back to the context and reminds us that there are social actors with stakes on this problem: organisers on the one side, and the Regional Government on the other. The reader should know that in Spanish, the adjective *justo/justa* is used both to signify 'just' and 'fair':

CARLA: But, let us say I am the Regional Government and I do not want to be just. (...) [If] I do not want to be accused of making up information, which mathematically correct way serves to be fair - not "just" [*does air quotes*] as in philosophical justice, but fair in the mathematical sense? (...) How can I pick frames being fair, mathematically?

BASTIAN: At random!

CARLA: [*Confirms*] At random.

In our planning, the reason for using random sampling is to fulfil the hypotheses of the Central Limit Theorem, which paves the ground for the normal confidence interval method that takes part in the official curriculum. The method is not mentioned until Session 2, so we push for students to come up with the strategies instead of giving them away. Interestingly, for students to include randomness as part of the sampling strategy, the notion of justice must come in. This idea is one of the most ancient functions attributed to chance (see Section 3.1); it transmitted the will of God and conveyed just decisions (David, 1962). Carla's strategy and students' response is an illustrative example of the insistence of researchers to pinpoint the role of context in statistics education; it gives meaning to otherwise obscure mathematical constructs (G. W. Cobb & Moore, 1997; Moore & Cobb, 2000; Garfield & Ben-Zvi, 2007; Wild et al., 2018). In the context of the task, we do not want to lean towards the case of the organisers with high-density frames, nor benefit the version of the Regional Government by picking low-density frames. We want to be *just*, and the "mathematical" sense – as Carla puts it – of being just or fair is picking at random.

The one feature of a sample yet to uncover is how many should we sample. Though many students show an intuition of the kind *the more, the better*, there is no straight forward answer to how many frames they ought to consider for counting. Instead, we want to build on their intuition and dig into what it means that larger sample size is *better*, by exploring it empirically. From this point, students begin their independent hands-on work by collecting samples. They assign a number to each frame and use their scientific calculators to generate random numbers. For each sample; they count the number of people on each frame of the sample and compute the sample mean of people. The third page of their worksheets has a table to register these three measures: sample size (n), number of people $(x_1, x_2, \ldots, x_n)$ and sample mean $(\overline{x})$. The goal of this activity is to collect data to address the matter of the influence of sample size on our estimation strategy; first by conjecturing and then contrasting with their generated data.

8.1.3 Making sense of statistics

During the first session, as a class we have set the scene and defined the problem, discussed an overall plan for making an estimate, and collected data in the form of samples; corresponding to the first three steps of the investigative cycle in statistical inquiry (Wild & Pfannkuch, 1999). As teacher and researcher, our planning follows specific curricular coverage and time restrictions. Still, the approach is to have students discover or negotiate their strategies through dialogues rooted in a context.

At this stage, students are using statistical language in conversations regarding the problem and the production of data. The problem, based on the news articles, is already stated in terms of *estimates*. The construction of the grid lets them face *variability* within frames. The strategy calls for taking *samples*, which should be taken *randomly* for the sake of fairness between parties involved.

I have mentioned that, in the analysis of the joint reflection in Session 3, the episode illustrating the divide between lies and statistics was coded in the *making sense* category (Weiland, 2017). In the same analysis, one more reflection is coded likewise, which refers back to our work on Session 1 as a whole:

RAIMUNDO: There was a hand... Was there a hand risen over here?

CAMILA: I think [the activity] was good, because it helped us, with the same content we are going through, to understand something that attracts the attention of us all.

I interpret Camila's comment as giving value to the efforts made by Carla and me to cover the same content they were supposed to study but anchored in a meaningful context. It was not easy, but the dialogical features of the session to build a random sampling strategy are evidence for it to be possible. Reformulating and challenging students and making use of resources such as the non-representativeness of a single frame to the whole picture, led to take a sample. The meaning of chance as fairness between opposing parties led to make the sample random. What is left to explore by the end of the session is the effect of sample size.

8.2 Session 2: Statistical estimation

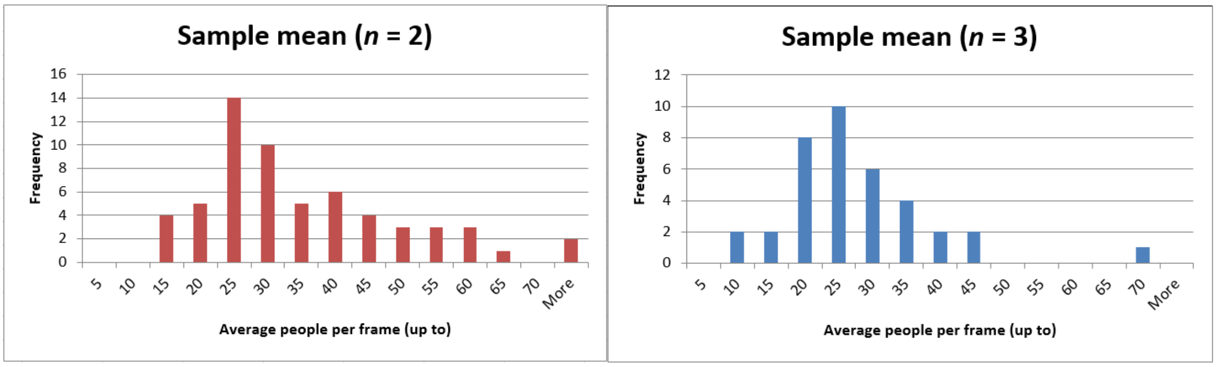
The inclusion of statistics in the curriculum in the last decades has grown not only at secondary and higher levels, but also in primary education. However, at lower educational levels, statistics contents tend to concentrate on descriptive statistics, computation of averages and interpretations of graphs. Statistics is portrayed in a "very narrow and limited way, which can be encapsulated: *every phenomenon can be captured by a bar chart*" (Ben-Zvi & Sharett-Amir, 2005, p. 1, emphasis in original). An obvious bottleneck towards its expansion is the level of mathematical content that primary students can make use of for formal statistical inference or modelling, though even professionals have trouble using and interpreting inference methods properly (Erickson, 2006). In the case of inferential statistics, a research trend has developed a broader notion of *informal statistical inference* (Makar & Ben-Zvi, 2011; Makar & Rubin, 2018), conceiving inference as a type of thinking or reasoning with a particular purpose, beyond its formal procedures.

Within this young tradition, a working definition for inferential reasoning is "the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (Zieffler et al., 2008, p. 44). For Makar and Rubin, this loose definition of inference as a probabilistic generalisation from data can be framed in three principles: (1) articulating the uncertainty embedded in an inference, (2) making a claim about the aggregate that goes beyond the data, (3) being explicit about the evidence used. The point is not to make a definition in a strict sense of the word; on the contrary, the goal is to "broaden accessibility to inferential reasoning with data" (Makar & Rubin, 2009, p. 85).

By the end of the first session, students had constructed and collected data in the form of random samples, following the third principle. The task is designed according to the curriculum, whereby the teacher ought to cover the notion of sampling distributions and normal confidence intervals, mediated by the Central Limit Theorem. As being a 12th-grade class, these contents can help to bring a richer language to articulate the uncertainty in terms of probabilities. They can also express an estimation in the form of an interval. However, the reader ought to know that this group of students had not yet studied the normal distribution, let alone the Central Limit Theorem, at the time of the classroom experience.

8.2.1 An authenticity dilemma?

Given the time limitations, students were encouraged to take samples of size 2 and 3 by the end of Session 1. Despite the variation encountered in their constructions of the grid (see Section 8.1), we agreed with Carla, in our meeting in between sessions, to gather all of them and display the results into histograms. I aggregated and produced the histograms seen in Figure 8.6. By examining back the students' worksheets, two of the three outliers (sample means above 70 people per frame) correspond to a 6×3 grid as seen in Figure 8.5-(c).



(a) (b)

Figure 8.6: Histograms of sample means for sample sizes 2 (a) and 3 (b).

The point is to pave the way for the normal confidence intervals by illustrating a particular attribute of the normal distribution. That is,

Theorem 8.1 *if* $\{X_k\}_{k\in\{1,\ldots,n\}}$ *is a succession of independent equally distributed random variables with* $X_k \sim \mathcal{N}(\mu,\sigma), \forall k \in \{1,\ldots,n\}$*, then the sample mean* $\frac{\sum_{k=1}^{n} X_k}{n} = \overline{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$.

As the sample size grows, the distribution of the sample means remains Gaussian, preserves the mean and reduces the standard deviation in inverse proportion to $\sqrt{n}$. The intuition we want to build from the histograms is that sample size preserves shape and centre, while it reduces spread. The strategy, guided by Carla, is to identify the common and distinctive attributes between the histograms in Figure 8.6, which is projected onto the whiteboard:

CARLA: Can you spot any difference or similarity?
CESAR: [*Points to (a)*] 25 is the largest frequency.
BASTIAN: [*Shows off*] The mode.
CARLA: But remember that is not a number but an interval (...) so it is between 20 and 25.
RAIMUNDO: [*Clarifying*] Between 20 and 25 is the most repeated average in samples of size 2. How about size 3, now that we are at it?
CESAR: That is what I meant. That is a similarity. They have the same mode.
CARLA: Yes, CÉSAR, I agree with you.

Students are accurate about what exactly can be drawn from the histograms; the mode. Histogram (a) is not centred enough to make the connection to the mean and the outliers do not help either. After a few minutes of asking about the shape, Carla decides to take the visual aid one step further:

CARLA: I know this is difficult (...) I want you to *erase* this one right here [*points to outliers on histogram (a)*]. Erase it from your minds for a bit. Now I want you to describe the graph without it. Can you spot anything special?
DANIELA: An increase and a decrease [*does hand gesture of a bell-shape*] (...)
CARLA: And on this one [*points to histogram (b)*], does the same happen? Exactly the same contour?
DANIELA: No, not exactly.
CARLA: OK, DANIELA, I am going to ask you a favour, to come and draw the contour above [histogram (a)] as you said, and I will make an effort to replicate the same curve up there [histogram (b)].
[DANIELA *and* CARLA *draw bell-shaped curves above the bars on the histogram*]
CARLA: Let us compare these curves (...) Which line is longer?
ALL: The orange one (on histogram (b)).
CARLA: (...) What does it say about data, how are they among themselves?
EDUARDO: More disperse.
CARLA: And what does it mean, EDUARDO?
EDUARDO: They are distributed further away from the average.
CARLA: And do you remember how is that called?
EDUARDO: Standard deviation... or variance!

We have no solid argument to claim that the number of people per frames is normally distributed, but students visualise the bell shapes easily as their hand gestures indicate. A benefit of collaborative work is that a large number of samples collected wager in favour of convergence. The histograms are constructed with 60 samples of size 2 and 37 samples of size 3. A frequentist approach to probability (Batanero, Henry, & Parzysz, 2005; Batanero & Sanchez, 2005; Borovcnik & Kapadia, 2014) asserts that the empirical distribution of repeated instances of a chance experiment would resemble its theoretical *a priori* distribution.

In our investigation, histograms resulting from data collection are not symmetric, and their dispersions are not so different from the blind eye. However, no all the blame is to be put on the unpredictability in data collection. According to Theorem 8.1, if X is the random variable representing the number of people per frame and it is normally distributed with mean μ and standard deviation σ, then the sample means for sample sizes 2 and 3, respectively σ_2 and σ_3, would be in the ratio:

$$\frac{\sigma_2}{\sigma_3} = \frac{\sigma/\sqrt{2}}{\sigma/\sqrt{3}} = \frac{\sqrt{3}}{\sqrt{2}} \approx 1.2247.$$

Therefore, the dispersion of sample means distributions are, by design, not different. In fact, by examining students samples, the ratio between empirical standard deviations s_2 and s_3 is $\frac{s_2}{s_3} = \frac{17.904}{10.820} \approx 1.655$ and it results in 1.566 disregarding outliers. That is, the resulting collection of data was, in fact, generous for our purposes, by displaying a more accentuated difference in the spread than the one we would have expected theoretically.

Carla addresses the issue of the centre by ignoring outliers that may be distracting, which, in hindsight, is justifiable after seeing which constructions led to such large sample means. Moreover, had we realised this between sessions, the matter of outliers could have been discussed transparently with students. The issue of spread requires extra visual aid, sketching contouring lines over the histograms. It allows Carla to exaggerate histogram (b) to make it look pointier than histogram (a). After these curves are displayed, not only students can spot the difference, but one (Eduardo) uses prior knowledge to identify two measures of dispersion, namely standard deviation and variance. A simple solution for future applications is to compare samplings of sizes that differ more than 2 and 3.

The transcript above, in reality, contains several pauses in between. This episode illustrates some of the challenges of using real data (from its source to its collection) to provoke insight into statistical ideas. Though researchers advocate for students producing and collecting data to address inquiries, a tension arises by the fact that "features of the resulting data sets cannot be predicted" (Ben-Zvi, Gravemeijer, & Ainley, 2018, p. 492). A possible solution is to imagine and tailor data sets to focus on the statistical ideas aimed to be put forward (e.g. P. Cobb et al., 2003). In our task, such an approach would miss the point of students' independent work on the previous session.

Moreover, the students' engagement and the relevance of the task are tightly anchored to its authenticity. Two lessons connected to the dilemma of authenticity arise from this episode. First, to make use of students' collaborative work to enrich the data set, in favour of attributes that depend on convergence in a frequentist meaning of probability. The second is to plan for construction and collection of data expected to illustrate contrasts that result evident.

8.2.2 Mathematical goals

Another well-known dilemma in didactical design is the one Ainley, Pratt, and Hansen (2006) call the "planning paradox", by which the design of a task is tensioned between engaging students by letting them use their own ideas to solve a problem, and the institutional instruction context obliging the covering of specific mathematical or statistical ideas. Ainley and Pratt (2014) propose to bridge this tension employing two complementing principles that ought to be in balance: a clear purpose for the task and the utility of statistical ideas in place. In the language used in this thesis, that is the tension between the *inquiry nature* of the learning environment and the *pragmatic* adherence to the curricular framework.

That is the case of the use of a normal confidence interval to make an inference. The study plan for 12th grade (MINEDUC, 2015a) only includes confidence intervals when the standard deviation σ is known and not estimated. Accordingly, we introduced the confidence interval that is constructed from a normal distribution, which is based upon the assumption that the sample mean is normally distributed, with $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. This is true if $X \sim \mathcal{N}(\mu, \sigma)$ (Theorem 8.1) or, by means of the Central Limit Theorem , if n is large enough with X having a population mean of μ and finite standard deviation σ. From here, the confidence 95% interval for the population mean can be derived by acknowledging that if $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$, then:

$$\begin{aligned} &\mathbb{P}\left(-z_{0.975} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{0.975}\right) = 0.95 \\ \Leftrightarrow\ &\mathbb{P}\left(-z_{0.975}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{0.975}\frac{\sigma}{\sqrt{n}}\right) = 0.95 \\ \Leftrightarrow\ &\mathbb{P}\left(\overline{X} - z_{0.975}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{0.975}\frac{\sigma}{\sqrt{n}}\right) = 0.95 \end{aligned} \tag{8.1}$$

For a random variable $Z \sim \mathcal{N}(0, 1)$, $z_{0.975}$ is the normal score defined such that $\mathbb{P}(Z \leq z_{0.975}) = \Phi(z_{0.975}) = 0.975$. Equation 8.1 makes use of the symmetry of the normal distribution around its mean, so $\mathbb{P}(z_{0.025} < Z < z_{0.975}) = 0.95$ and $z_{0.025} = -z_{0.975}$. From a pragmatic point of view, constructing Equation 8.1 has two big challenges, if we are to continue our dialogical and inquiry-based approach. We touched this issue during the interview:

CARLA: I think, in rigour, everything I say is mathematically correct. For sure. Still, one omits information.

RAIMUNDO: Yes. I mean, I remember very well that we *passed them two goals*. One was that the confidence intervals, at least those in the curriculum...

CARLA: ... are with known sigma (σ), right? And we had no way of knowing...

RAIMUNDO: Exactly. It is only an estimated sigma (s). That was one. The other is about the normal distribution. At least for the sake of this calculation, we used tables only. So we did not get into it so much and used the values from the table.

CARLA: You mean for the z [score]?

RAIMUNDO: The z, yes.

CARLA: (...) Oh, so we had not seen the normal [distribution] yet. Because I remember another time where we taught them how to calculate that level of confidence.

In Chile, *passing goals* is an idiom expression used as *fooling*, as an analogy to make an unnoticed goal in a football game. In this episode of the interview, I use it to make the case of two aspects of the confidence interval that we were not transparent about during the experience and which Carla remembered well. First, our standard deviation s is an estimate of an unknown population standard deviation σ. However, to deal with unknown standard deviations correctly, we would have had to introduce Student's t-based confidence intervals, which is beyond the reach of the curricular framework. The second problem is related to the students' previous knowledge, as they are not familiarised with the normal distribution at this point, so asking them to deduce and find the multiplier $z_{0.975}$ would have been a long shot. That is not to say that we ought to give up the use of normal confidence interval, but instead to present its calculation matching their intuitions. The following excerpt from session 2 reveals Carla's strategy to do so, by challenging students' baseless estimates:

DANIELA: [*Joking*] It is known that there were 40,000 (people).

CARLA: OK, let us see... Exactly 40,000?

DANIELA: But... it could also vary.

CARLA: OK, it could vary a bit. Yes, OK.

DANIELA: But close to 40,000 [*does hand gesture spreading thumb and index fingers*].

CARLA: But tell me... close. How do you make sense of *close*? You made that gesture, what does it mean?

DANIELA: It means it could vary. Not like close to 50,000, but...

CARLA: But... go on, go on.

DANIELA: I do not know.

EDUARDO: I say it is like an average between both, i.e. 45,000.

CARLA: Exactly 45,000? (...)

EDUARDO: Yes. No... [*makes a hand gesture of a central vertical line and an inverted 'V' from it*].

CARLA: Let us see...

RAIMUNDO: They are making another technical gesture.

CARLA: Those are the gestures I love, but you have to verbalise them because they are very...

DANIELA: Oh! Aprox[imate]!

CARLA: OK, "approximate". I do not like that word so much, because, for example, if he got a 6.8 (grade) and I say "approximately a 7.0", it does not work[16]. What is it you [*to Eduardo*] are doing with your hands and Daniela is doing with her fingers?

EDUARDO: It is like the *bell-shape of the histograms*.

CARLA: What is it that you just mentioned? Eduardo, say it aloud because I cannot hear you. (...)

DANIELA: [*Eureka moment*] I think it will be an *interval* between 40,000 and 45,000.

Carla's strategy is based on crucial dialogical features as described by Alrø and Skovsmose (2002). *Getting in contact* is first happening by the use of humour by Daniela at the start, and it is a good sign of a safe environment for discussion, and by asking Eduardo to speak aloud. *Challenging* is put to evidence by Carla's insistence on asking whether the students' estimates are exact. Most relevant for this episode is how she *identifies* informal ideas like "approximate", "bell-shape" and hand gestures, which she uses to *crystallise a mathematical idea*: making an inference in the form of an *interval*, based on the *bell-shapes* of the histograms.

During the interview, we discuss this episode as being worth of admiration and helpful for the inquiry approach. Carla claims it is institutionalised:

CARLA: That *style* is commonplace here. If you watch any colleague from this school, that is the style. [The style] of not giving away the answer, to take their answers, use them, take them. Even if they are incorrect, one can take them as a trampoline to reach the right one (...)

CARLA: And, in reality, you can take advantage of any answer, except for "I do not know". You just have to be patient, but it is very much the style of the school.

Right after Daniela's Eureka moment, Carla presents the formula for computing a 95% confidence interval. The confidence interval is given as:

$$\mu \in \left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}} \right). \tag{8.2}$$

Carla connects each of the elements in Equation 8.2 to intuitions already built in the previous discussions. The sample mean $\overline{X}$ lies at the centre of the interval, but since it is an approximation, the interval has a certain width. It should account for the standard deviation of the sample, which is given by s (the first goal we pass). As students are experiencing, the larger the sample, the histograms shrink, and thus σ is divided by $\sqrt{n}$. Finally, a similar dialogue as the one above leads to the intuition of *confidence*, whereby the only way of reaching 100% is to have an infinitely wide interval, and the more we shrink it, less confident we are. Carla briefly explains that for a 95% confidence, the bell-shaped normal distribution gives us the multiplier $z_{0.975} = 1.96$ (the second goal we pass).

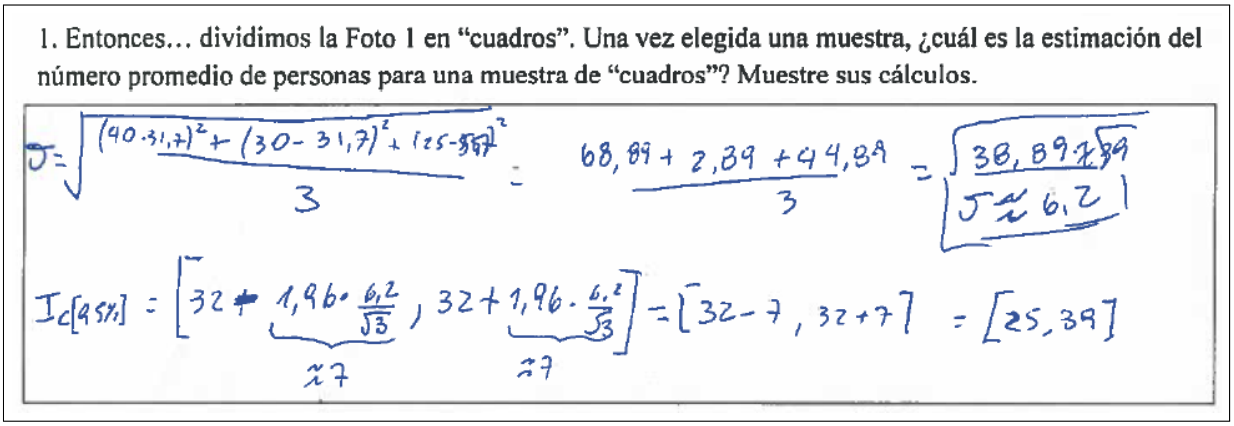

1. Entonces... dividimos la Foto 1 en "cuadros". Una vez elegida una muestra, ¿cuál es la estimación del número promedio de personas para una muestra de "cuadros"? Muestre sus cálculos.

Figure 8.7: Single frame 95% confidence interval by Roberto

From this point, students use their collected samples to compute their 95% confidence intervals for the population mean. For example, in Figure 8.7, Roberto computes a confidence interval based on a sample of 3 frames, giving that $\mu \in (25, 39)$. I recognise that we missed several mathematical underpinnings. For time's sake (*pragmatism* again), we did not deal with general problems interpreting confidence intervals. Their frequentist interpretation relies on repeating the sampling experiment and the construction of the interval itself (White & Gorard, 2017; Borovcnik, 2017). However, the procedure joined many of the intuitions built by the students' discussion from their collected data. These ideas include the shape and spread of histograms based on the sample size and the approximation via intervals. Still, calculations are necessary to follow the general inquiry. The confidence interval must be expanded to the full reach of the march. Only after that, they can, once and for all, answer how many people attended the students' march, approximately. This way, we pass them a couple of mathematical goals, in order to achieve mathematical goals (pun much intended).

8.2.3 The full extent

From my own collection of photographs from the media, Figure 8.3 was the most overreaching of all, and thus intended to scale up students' estimates to the full demonstration. However, in the process of designing the learning environment, Carla had me know that the march was larger than what is there depicted. She knew; she was there too. We decided to keep the picture as part of the task but printed out copies of a map of Santiago as in Figure 8.8, waiting for students to figure out and be prepared. For reference, Figure 8.3 fits upside down in the rectangle around Baquedano Metro Station in Figure 8.8. It did not take long for them to realise once they had finished estimating the number of people in Figure 8.2, i.e. by multiplying both extremes of their confidence interval by the number of frames.

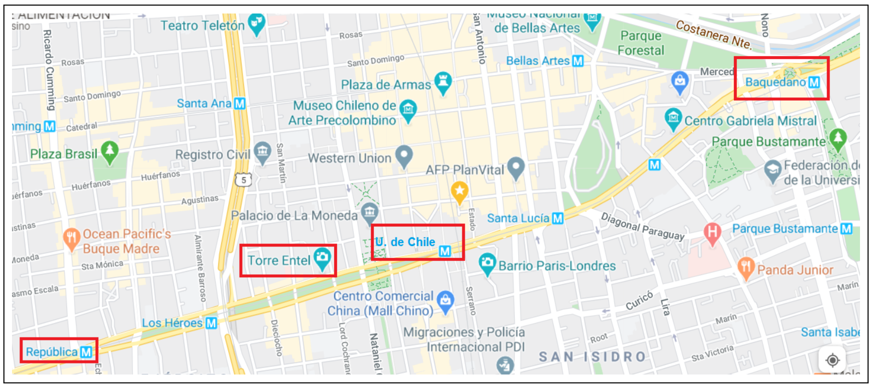


Figure 8.8: Map of Santiago centre used by students.

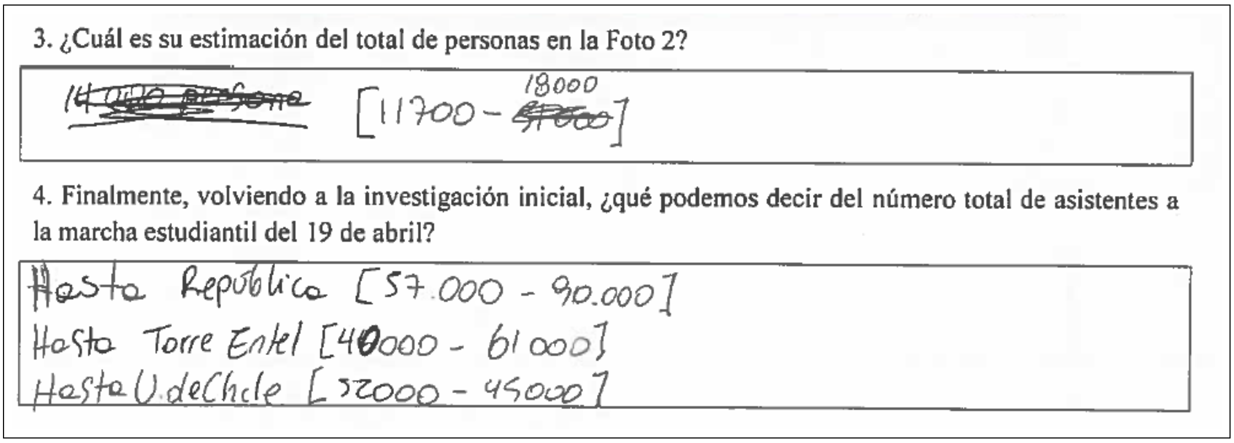
The question is then, for how much to multiply their new interval to extend to Figure 8.3 or some extent in Figure 8.8. The answer depends, evidently, on the assumption of extension of the march. After some discussion, Roberto’s group finds an interesting strategy to compute estimates according to different assumptions as seen in Figure 8.9. Their 95% interval for the mean of people in a single frame (Figure 8.7) is multiplied by 36 to estimate the number of people on Figure 8.2, that is $(36 \cdot 25, 36 \cdot 39) = (900, 1404)$. Based on their assumptions, the trapeze constructed on Figure 8.2 fits 13 times in Figure 8.3, thus giving an interval of $(13 \times 11,700; 13 \times 18,252) \approx (11,700; 18,000)$. The next step they take is to make three different extensions, depending on whether the march reached to three different spots. These locations are República Metro Station (overestimate), Torre Entel (middle case), and Universidad de Chile Metro Station (underestimate), multiplying (and approximating) their interval for Figure 8.3 by 5, 3.4 and 2.5, respectively.

In the analysis, amongst the categories of *writing the world* of critical statistical literacy, one episode during the joint reflection is illustrative of “negotiating societal dialectical tensions when formulating statistical questions, data collection and analysis methods and highlighting such tensions in the results of a statistical investigation” (Weiland, 2017, p. 41). This episode referred precisely to their strategies for extending their estimate to the whole extent of the march, right after they explained why Figure 8.3 was not enough:

SILVIA: Yes, supposedly the march reached further, but not with the same mass, that is, “relatively” [*does air quotes gesture*] with the same mass. As if they were moving all together with the same people.

ANIBAL: But we also calculated until Torre Entel. More or less, we still assume that all that complete section [is] full of people. So we still calculate completely. [*To* EDUARDO] How much did it gave?

EDUARDO: Until Torre Entel, it gave between 40,000 and 61,000.



3. ¿Cuál es su estimación del total de personas en la Foto 2?

[11700 - 18000]

4. Finalmente, volviendo a la investigación inicial, ¿qué podemos decir del número total de asistentes a la marcha estudiantil del 19 de abril?

Hasta República [57.000 - 90.000]
Hasta Torre Entel [40000 - 61000]
Hasta U. de Chile [32000 - 45000]

Figure 8.9: Full extent intervals by Roberto

RAIMUNDO: Between 40,000 and 61,000.
ANIBAL: So we still believe it is quite high.

My expectation for this part of the activity was that students would blandly overestimate the number of attendees. However, this episode suggests that they are aware of dialectical tensions between interested parties and their willingness to underestimate or overestimate. As a point of departure, their attendance to the march makes them realise that Figure 8.3 is underestimating the reach of the march. They know the march reached República and computed accordingly. However, Silvia highlights the fact that the march is *moving* from Baquedano to República, so it would be an overestimate. In response, the group was prepared to bring balance and suggest a middle ground estimate. This estimate is higher than the authorities', so they "still believe it is quite high".

Session 2 of this classroom experience made us face several challenges, mainly regarding the tension between a *pragmatic* adherence to the curricular framework and the *inquiry approach* to students' work. In order to achieve our mathematical goals of making an estimate employing a confidence interval, we had to pass a few mathematical goals and make Equation 8.2 available to students. This approach differs from the previous chapters. Our co-flections (collective reflections, see Valero (1999)) during the prior meetings and those in-between sessions were a necessary input for agreeing on a strategy, highlighting the negotiation aspect of *practical organisation* (Skovsmose & Borba, 2004). The balance was reached by trustful collaborative planning with Carla, illustrated by our transparent conversations during the interview. More importantly, it was possible by letting her do her art in the classroom. Carla's dialogical style of teaching turned students' discoveries and reflections into an established mathematical tool. It bridges the second and third stages of the inquiry approach described by Blomhøj (2016).

Many mathematical flaws and holes are not addressed in formal terms. However, the instrumental exemplarity is experienced in the task, for the learning environment is exemplary for the big idea of statistical estimation. If we come back to the framework for informal inference, suggested by Makar and Rubin (2009): (1) The collection of random samples and the construction of 95% confidence intervals account for the uncertainty embedded in the inference. (2)

The extension of estimates from the sample of frames to the bigger pictures is made under explicit assumptions of generalisation and full reach. (3) The calculations are based on collected evidence from actual pictures of a real event.

8.3 Session 3: Casting shadows over sources of data

Let us take a closer look into one of the categories of critical statistical literacy belonging to *reading the world*, namely the one defined as "evaluating the source, collection and reporting of statistical information and how they are influenced by the author's social position, and sociopolitical and historical lens" (Weiland, 2017, p. 41). Though the notion of statistical information can be broad; source, collection and reporting describe different activities (Wild & Pfannkuch, 1999). In the context of the task, the source and collection come together in the form of aerial pictures. The reporting is what is read on the media, that is, 30 thousand people attended the march, according to the Regional Government, as the media outlet reports. This way, there is a distinction to make between data and interpretations/claims (Kuntze et al., 2017). From the analysis, utterances in the *evaluating* category come up both during the joint reflection by the students and during the interview with Carla, and they relate to a particular part of the discussion at the end involving rich students' context knowledge.

8.3.1 Students' context knowledge

The role of context is a vital issue when dealing with critical statistical literacy, as context gives meaning to data and statistical messages altogether. For this reason, context or world knowledge is, in fact, within the knowledge base proposed by Gal (2002), given that, "if a listener or reader is not familiar with a context in which data were gathered, it becomes more difficult to imagine (...) how a study could go wrong" (Gal, 2002, p. 15). The design of the task brings this idea to an extreme, as some students participated in the march. They have a well-developed context knowledge base; they are part of the context. This feature plays out as a ground for their final reflections.

During the second half of the third session, representatives of each group of students are asked to share their results. As discussed in Section 8.2, students realise that Figure 8.3 does not cover the whole march, grounded by the fact that many of them were there. This recognition led to different assumptions on how to expand from Figures 8.2 and 8.3 to some extent of the map on Figure 8.8 and consequently adding up to the variability amongst the students' estimates of the total crowd. The discussion about the methods and assumptions triggered comments about the data source itself, as reflected on the following excerpt:

RAIMUNDO: You attended the march.

ANIBAL: Yes.

ROBERTO: The march reached a bit further. So that is why I... [EDUARDO *interrupts*] Because it was not only that... [SILVIA *interrupts*] But maybe this picture was early. That is what... [*even* RAIMUNDO *interrupts*].

RAIMUNDO: And that is interesting, but here – which you also commented about – you were in place.

...

SILVIA: Yes, supposedly the march reached further, but not with the same mass, that is, "relatively" [*does air quotes gesture*] with the same mass. As if they were moving all together with the same people.

...

ANDRES: I mean, it is just that the (...) path of the march goes from Baquedano until República (see Figure 8.8), not that the march extends all along, (...) that is, the march would be moving along that.

One of the aspects standing out from this dialogue is that evaluating features of the sources came without an explicit prompt. If any, my prompt was pointing – insistently – to the fact that some of them were actually at the event. While making a case for the further reach of the demonstration, Roberto – despite interruptions – pointed out that the pictures may have been taken early. Other classmates bring up the fact that partaking in a march implies moving along a path, and not merely staying in one place.

The other aspect worth highlighting is how detailed their argumentation becomes once we let them unfold their context knowledge. Alrø and Skovsmose (2002) call this dialogical feature *identifying*, enabled via *why-questions* which call for *justifying* an interlocutor's point. The discussion goes on, and my role as mediator becomes more active, that is, I ask students to follow up on and explain the points they are making:

EDUARDO: I think that [photo of the] march was early. Why? Because it (the street) was not even closed on the other side of Alameda [avenue].

ROBERTO: In fact, there are still buses.

...

RAIMUNDO: Why did you mention the buses? Why is it relevant that there are buses?

ROBERTO: Because at that time, after an hour, they block the transit, and no buses pass by. So, in this photo there are buses, it means it was too early.

BASTIAN: The march had not yet begun.

...

RAIMUNDO: There was a [raised] hand...

ANDRES: Oh, I wanted to say that ROBERTO must be right. That about the [photo of the] march being early, because it is like, observing more, yeah, but if it is known that the north is down [in the picture] and the shadows are from there [*points left/East*], because we are super early, before 12.

RAIMUNDO: Right.

ANDRES: And the shadows of the buildings [*general laugh follows*].

RAIMUNDO: All right, that reasoning... [*class too noisy.*]

CARLA: Class, let us focus, please!

There is a clear step up from interrupting the Roberto to asking him to build on his mention to buses. I do not give up the discussion, but instead, I let the student raising his hand make the case of the shadows. Moreover, after the general laugh at his comment, both Carla and I reassured Andrés that his comment was utmost on point, as she reflects:

CARLA: Oh, because they talked about the shadows, yes. I remember I saw it thereabout.

RAIMUNDO: Because then it was ANDRÉS who insisted on the shadows and I realised now, that I watched it again, that a majority laughed, as in "you are exaggerating" about the shadows of the buildings.

CARLA: [But] it made much sense!

A variety of arguments for how early the pictures were taken comes up. A street, which they knew was closed at some point, is open in the picture. The pictures display buses, which stop running once the marching begins. And so come the shadows, their direction and length indicate the pictures were taken early in the morning. Moreover, almost two minutes after the discussion above, Aníbal said to his group "Yeah, not even the *guanaco*[17] was wandering around here", which he knows should be in place when the demonstration is at its peak.

These exchanges illustrate the role context knowledge can play in evaluating the source of statistical information, in particular, when students involved in the inquiry have taken part in the contextualised event itself.

Coming back to the analysis on NVivo, this episode is coded not only as students evaluating the source of data, but also as one category of writing the world through statistics, namely "communicating one's social location, subjectivity, and political context to others and how it shapes one's meaning-making of the world when reporting results of a statistical investigation" (Weiland, 2017, p. 41). The dialogical character in the classroom, manifested by letting students communicate their subjectivity (a *writing* category), enables them to evaluate the source of information (a *reading* category). This ability of the task is, of course, a particular feature of its design, namely that the principle of *exemplarity* is applied from the subjective relevance perspective (Dewey, 1938/2015; Illeris, 2002). They know the context because they are part of the context:

CARLA: Well, it looks like they know better than us that the march goes further from one side to another and at a reasonable time, (...) let us say, three in the afternoon.

8.3.2 Dynamics and timing

By reacting to the classroom episode above, Carla was able to identify two main themes in the students' appraisal of the pictures as valid sources of data:

CARLA: I mean, I see two things here. (...) The one about them managing to perceive that it (the march) is not something constant; that area may not be completely filled. Another thing is that they manage to perceive that if it was earlier, there are fewer people, and then more people accumulate.

The pictures do no account for the dynamic nature of a march that is in constant motion, making the expansion from one picture to the full reach of the demonstration problematic and ambiguous. The time where they were taken is too early to represent the extent of the attendees, so any estimate made from them can be deemed an underestimate. Carla's suggestion is to find a way to obtain pictures of the demonstration at an hour where it is at its apogee, for example, at three in the afternoon. My impression was slightly more negative as if the task itself fell like a house of cards.

RAIMUNDO: ...what I highlight from what the kids say is a criticism to the activity itself. Because the estimation methods we were using, using the pictures, make sense only if this was static. So maybe this method or the context was not appropriate for the activity. At most, it is appropriate in the sense of allowing them to discuss it.

CARLA: I think there lies the potential of this. After all, if it were static, there would be no point whatsoever. Because this is moving in time. So, perhaps, what they could have asked themselves is "what would have happened if I take a picture at this hour and move, say, two blocks?" Considering time and distance, one can see and compare or even reconstruct from several photographs.

We agree on the technological knowing (Skovsmose, 1994), observed by a technological discussion (Barbosa, 2006) that is taking place, whereby students question the appropriateness or reliability of the mathematical approach used for the purpose in mind. However, here we find an instance of how the *instrumental exemplarity* of the task is defined in a dialectical manner. On the one hand, dynamics and timing make the task invalid concerning its instrumental exemplarity (Wagenschein, 1956; Illeris, 2002) since the methods of estimation are not appropriate in the context of the march. On the other hand, it is still exemplary for inference and estimation (Zieffler et al., 2008; Makar & Rubin, 2009) as a statistical *big ideas* (Garfield & Ben-Zvi, 2008; Ben-Zvi, Gravemeijer, & Ainley, 2018).

Carla has a more radical approach. Not only she accepts the problem of dynamics and timing, but she believes it is the whole point of doing the investigation in the first place. She argues for this potential by imagining new possibilities in the form of constructed episodes (Blomhøj, 2006), that is, imagining dialogues that *could* have taken place, anchored in the outcomes of the task and her familiarity with her students. From her point of view, by identifying the inappropriateness of the methods intended, the task needs not to be discarded, but opened up and upgraded. She proposes to go further and have students investigate on the internet and even get out of the classroom and ask authorities about their methods.

Carla's pedagogical imagination (Skovsmose & Borba, 2004) may enable students to engage from technological to reflective knowing, as the first entry point of such process is to ask whether we need in fact need mathematics (Skovsmose, 1994). In this task, the question can be rephrased as "do we, in fact, need statistical estimation?" Here, two of the design principles clash. In order to push for the open-ended inquiry nature of the task, pragmatic attachment to the curriculum is compromised. As Carla puts it:

CARLA: But that is still a path to investigate. Perhaps, one may "lose" a little bit of mathematics. I mean it in between double quotes, because investigating is part of the scientific, so it would still be useful.

From this perspective, the mathematical approach should be at the service of the investigation and not the other way around. Her take on the design of learning environments is to shift from the paradigm of exercises to landscapes of investigation (Alrø & Skovsmose, 2002; Skovsmose, 2011).

There is a double reading for the title of this section. For once, there are shadows literally casting over the streets captured by the aerial pictures, which work as data sources for the task, brought up by the students' context knowledge. At the same time, these shadows enable to challenge two features in the task; the validity of the pictures as sources of data and the methods as necessary for addressing the inquiry. Carla and I recognise the flaws in data sources. She uses her pedagogical imagination to find new possibilities of opening up the investigation, while I revise some of the theoretical conflicts between design principles. All of this was possible because of students' reflections casting shadows over the contextualised task.

8.4 Critical statistical literacy

In this chapter, I have described several features of a statistical estimation task that balances three perspectives on *exemplarity*. First, it is anchored in a real event that many students experienced themselves, namely a students' march, allowing them to enrich the discussion based on their context knowledge. Second, it is exemplary for the big statistical idea of inferential reasoning (Zieffler et al., 2008; Makar & Rubin, 2009). Third, it is exemplary of the way crowd size estimations are entangled with political motives to steer the public's adherence to an agenda. To a large extent, the balance in design and implementation of the learning environment was achieved by a trustful communication with Carla, the teacher, living up to the principle of *pragmatism* as a transparent negotiation.

Throughout the chapter, I made use of a framework for analysis that is constructed as an intersection between notions of critical literacy and statistical literacy (Weiland, 2017). Though many statistical literacy frameworks exist in the literature (e.g. Gal, 2002; Watson & Callingham, 2003; Sharma, 2017), the choice was meant for exploring a possible connection between domain-specific ideas and those of critical mathematics education. However, I could pinpoint three main issues of this choice.

8.4.1 High expectations

The descriptions of categories are very ambitious, so, in order to get the coding to work, my criteria had to be relaxed. For example, the episode of students *negotiating* assumptions to expand their estimates to the full reach of the march was coded as "negotiating societal dialectical tensions when formulating statistical questions, data collection and analysis methods and highlighting such tensions in the results of a statistical investigation" (Weiland, 2017, p. 41). This coding is a noticeable stretch, considering that Weiland refers mainly to the operationalisation of variables in statistical investigations and the way their findings are reported. In part, it is related to the social construction of statistics pinpointed by Best (2005), but also to the political framing of redefining variables such as unemployment or poverty (Frankenstein, 1994). In the coded episode, the tension is political but tacit. Students are not explicit about how convenient it is for the government to underestimate and for organisers to overestimate the crowd size, in that episode. However, my coding is optimistic, spotting the potential to reflect upon the societal or political tension as a quality of the task.

Perhaps, the ambition of the framework's categories is commonplace to other statistical literacy frameworks as well. Batanero (2002) comments that statistical literacy should be seen as an ultimate goal and guidance for policy and curriculum. However, teaching and research at micro-level should be supported by local frameworks, as I did in previous chapters. After all, the expectations of a statistical literacy are too high to be the fruit of one classroom experience.

8.4.2 Reflective knowing

To a large extent, the framework does not intersect the version of *critical* I am using throughout the thesis. I take, like a red thread to address **RQ4**, the fifth entry point on reflective knowing (Skovsmose, 1992, 1994), by which students should reflect about how applications of mathematics affect our conception of some part of the world. However, the closest category in Weiland's framework is the one of "identifying and interrogating social structures which shape and are reinforced by data-based arguments" (Weiland, 2017, p. 41). It is quite close to identifying and interrogating the formatting power of statistics.

Revisiting the analysis in NVivo, students do engage in critical reflections within the *interrogating* category, intersected with *making sense*. Some of those utterances were mentioned above and include statements such as "the institution lies", "the more *state*-media lie" and "we cannot trust anyone; they take advantage of the situation; it is like a lie, but…disguised". I discussed this point related to overcoming the statistic-or-lie dichotomy towards a construction-and-calculation bundle (Best, 2005). However, from a CME perspective, students realise that mathematics (and statistics) is playing a role in disguising political messages with a veil of objectivity given by numbers. They are addressing statistics' formatting power.

The *interrogating* category takes part in the coding of the interview as well. Carla takes an attitude of projecting our experience into the future by proposing new relevant topics to include in her teaching practice:

CARLA: So we (the teachers) were thinking about what contingent topics, where there are problems, we could approach from mathematics. And, I do not know, we reached as a conclusion that we really should teach about AFPs (Pension Fund Managers)[18] Because it is something attractive nowadays. It is true that students are not workers, they have nothing to do with it, but they do hold the narrative of "Oh, those AFPs! Death to AFPs". They say those things even if they are not directly involved, but their parents are workers, and they may have a grandparent with a very undignified pension. And that is a very mathematical topic!

Carla and her colleagues are realising that mathematics plays a substantial role in our society. Some structural issues, such as our pension system, are built on mathematical models and algorithms which can be part of classroom discussion. In the language of (Skovsmose & Borba, 2004), Carla is taking the new current situation produced by our experienced arranged situation, and using her pedagogical imagination towards new educational practices.

From my standpoint, what is critical in this task is how mathematics partakes in shaping our perception of popularity or adherence to a cause. In consequence, it can affect our views on the cause itself. That is, a cause can benefit from showing how large of an adherence it gathers, or, on the contrary, the opposing party can profit from their invisibility. This strategy is known as mainstreaming, and it is a practice through which an often fringe social group manages to turn into a visible and normal part of society. Though different academic definitions can be found, mostly in gender studies (Daly, 2005), at least it has been shown to be transformative in advancing the acceptance of diversity, by pushing policies in favour of people that once were deemed to be invisible (Squires, 2005).

I try to make this point by asking directly:

RAIMUNDO: And, besides, for me it was quite important to know whether these numbers influence you. For example, your classmate already said this produces him a certain opinion about the media. But now that you have these numbers, what... do they produce anything in you? Do they change any of your opinions about the march, about its causes?

Students do not react in the direction I wanted to. Instead, they change the topic towards statistical "lies". This shows how a direct question will not necessarily provoke critical utterances if not framed correctly within the task.

8.4.3 Reading and writing

Weiland's notion of critical literacy is based on the work of Gutstein (2006) and others into the relationship between mathematics education and social justice, understanding literacy in an emancipatory sense inspired by Freire (1970/1993). Weiland (2017) advocates for both understanding structures of social injustice and using statistics to change them. Mathematics education for social justice and the formatting power of mathematics are different preoccupations of the CME agenda (Skovsmose, 2014a; Valero et al., 2015), which explains the point I made above about addressing **RQ4**.

However, the social justice inspiration conveys a relevant feature of Weiland's framework, namely the distinction between reading and writing the world with statistics. In previous chapters, the tasks' ability to provoke critical reflections from students was discussed in terms of the more or less active role of students in the inquiry work. In this chapter, the analysis allowed me to split waters between critiquing statistical information or data-based arguments that are consumed by students (Gal, 2002), and the process of critique that takes place rooted in students' own statistical inquiries (Wild & Pfannkuch, 1999). Moreover, the overlaps suggest that students' inquiry work enables them to be critical of others', echoing the calls for venturing into landscapes of investigation or inquiry-based practices as a way to connect mathematics education, democracy and citizenship (Alrø & Skovsmose, 2002; Artigue & Blomhøj, 2013; Skovsmose, 2011).

8.4.4 Coda: Writing the world

At the beginning of this chapter, I pinpointed that the analysis did not reveal any utterance coded as *alleviating*, within the framework of critical statistical literacy by Weiland (2017), as the learning environment does not provide opportunities to take actions beyond the classroom setting. There is already potential to write the world with statistics by using statistical investigations, negotiating dialectical tensions and communicating one's subjectivity. However, what could students do in a statistics learning environment to resolve sociopolitical issues of injustice?

Part of the motivation of the task was the fact that regional authorities give crowd size estimates in a non-transparent way and they tend to be taken at face value, especially if the numbers seem reasonable, as it is the case of the march on April 19th, 2018. However, a turning point occurred after the Women's March on March 8th, 2020. The police made an official estimate of 125,000 attendees, which was suspiciously low for organisers and participants. This time, other associations made their estimates. An academic from the Faculty of Architecture, Design and Urban Studies of the Pontifical Catholic University estimated at least 800,000 attendees and made a full display of his methods in the press (Castillo & Labrín, 2020). An academic from the Geography department of the University Academy of Christian Humanism estimated between 1 and 1.5 million attendees. In her opinion, "the government yet again belittles the figures of attendees to these type of mobilisations. The government wrongfully measures things it finds inconvenient" (Manuschevich, 2020). After reactions, the police explained that their methods are based on drone pictures and mathematical models, although they did not make them transparent.

What is the sociopolitical issue at stake? Which dynamics of power provoked this turning point? During the social unrest at the end of 2019, Human Rights agencies denounced that the Chilean police force was systematically violating human rights, in some cases, sexually abusing women detained during the protests.[19] The feminist movement took a stand against the National Police Force. The most iconic example was a chant by the collective LASTESIS that spread around the world, where they criticise the role of the police in the systematic patriarchal structures of our society. As of the writing of this chapter, the Police Force is filing charges of intimidation against the feminist collective.

The students participating in the "How many people attended the students' march?" have stakes on the issue of the march, grounded in their attendance to it. The experience shows that, other than the access to quality data sources, there are no substantial obstacles for making their own estimates, discussing assumptions, and negotiating steps in the procedures. Perhaps, a step further would be to implement the learning environment encouraging students to raise their voice and publicly (e.g. as a letter to a newspaper) respond to the inquiry of how many people attended a students' march. Skovsmose (1998); Alrø and Skovsmose (2002) show that talking back to authority using mathematical investigations of public interest is a crucial connection between mathematics education and citizenship. I prompt this idea to Carla during the interview:

RAIMUNDO: But I do not know if students here, especially from 11th and 12th grade, have had these initiatives to disseminate or go out.

CARLA: No, here we are too fearful of going out of school. But it is nothing but apprehension with them. But they could do it. I think now, with the new curricular framework, it could be encouraged. Now we have a[n official] justification. It is like "now we have to do this".

Carla refers to the new curricular bases that begin to rule from 2020 (MINEDUC, 2017, 2019), which, as discussed in Section 3.3, represents students as social actors with higher protagonism. Taking their investigations outside of the classroom to write their world with statistics is possible.

Notes

[13]*Confederación de Estudiantes de Chile*, Chilean Students Confederation

[14]*Asamblea Coordinadora de Estudiantes Secundarios*, Coordinating Assembly of Secondary Students

[15]*Coordinadora Nacional de Estudiantes Secundarios*, National Coordinator of Secondary Students

[16]In Chile, grades are usually in a scale from 1.0 to 7.0, using one decimal position.

[17]*Guanaco* is the name of a species of camelids from the Andes. Known for their tendency to spit, it is also the folk name given to the police water cannon truck. The teargas truck's folk name is *zorrillo* (skunk), for obvious reasons.

[18]Chile's current pension system was installed in 1981 during the dictatorship, with a strong focus on individual capitalisation through Pension Fund Managers (AFP - Administradoras de Fondos de Pensión in Spanish). The resulting perpetuation of income inequalities in pensions (Comisión Presidencial Pensiones, 2015) have put those institutions at the core of current calls for reforms.

[19]See, for example, https://www.amnesty.org/es/latest/news/2019/11/chile-responsable-politica-deliberada-para-danar-manifestantes/

Chapter 9

Binding discussion

Before addressing the research questions and the coherence along the general *problématique*, my purpose in this chapter is to revise the four learning environments as a whole, by identifying relevant similarities and particularities, in light of the classroom experiences. First, I highlight relevant topics connected to the preoccupations of CME that were illustrated in the designs. The formatting power of mathematics lies at the centre of the thesis. Nonetheless, connections to other preoccupations emerge as well. Second, I discuss the research focus on reflective knowing (**RQ4**), and how it is intersected with the analyses, employing different domain-specific frameworks (**RQ5**). Third, I revise common tensions between the design principles as they developed during the interventions in order to redefine them and address **RQ3** later in the conclusion chapter.

9.1 Towards a critical justification for probability and statistics

As discussed in Chapter 3, the premise of critical justification for the learning and teaching modelling and applications of mathematics in school, as formulated by Blum and Niss (1991), includes the claim that the world's shape and functioning are (increasingly) influenced by mathematical models and applications. A necessary step towards a critical justification of probability and statistics in upper secondary school is to present ways in which probability and statistics, aside from other units of the mathematics curriculum, shape our perception and format automatic and deliberate action in the world. I now describe issues risen by the classroom experiences that give potential to the learning environments to illustrate the formatting power of probability and statistics, within the limitations of the school curriculum. These topics are the consequences of the assumption of independence, probability as a bearer of evidence, statistics as political arithmetic, and the mathematisation of risks. Moreover, other critical topics emerged during these experiences and, in retrospect, they could add to the critical justification as well, namely the notion of foregrounds, the construction of data, and the acknowledgement of decision stakeholders.

9.1.1 Consequences of the assumption of independence

Chapters 5 and 7 are exemplary for the role of the assumption of independence in computing probabilities of intersected events.

The fact that $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ is not a consequence of the independence of A and B but its definition, and therefore it can only be assumed or reconsidered in hindsight (Borovcnik, 2017). This reality is consistent with some textbooks on probability theory, where after the mathematical definition of independence is given, in terms of modelling, at most they can say, for example, that "in practice one usually has the correct *feeling* that certain events must be stochastically independent, or else the probabilistic model would be absurd" (Feller, 1968, p. 125, emphasis added). Statistical independence can be assessed, for example, using χ^2 tests of association, far from the reach of the upper secondary school curriculum. In some cases, it is claimed that "independence is often easy to grasp intuitively. For example, if the occurrence of two events is governed by *distinct and non-interacting physical processes*, such events will turn out to be independent" (Bertsekas & Tsitsiklis, 2000, p. 32, emphasis added). Independence is, therefore, contrary to association or interaction, which is often confused by causation. For example, in Chapter 5, students design the juice experiment so that one tasting trial does not affect the next one. They achieve this by having the subject wear a blindfold and not letting her know whether she guessed correctly or not. Later, when analysing the case of Malcolm Collins, students sustain that beard and moustache are independent, as they are different features. Certainly one does not *cause* or *affect* the other, but the real question is whether they are associated.

If acknowledged, a doubtful assumption of independence cannot always be amended with the use conditional probabilities to compute $\mathbb{P}(A \cap B)$, since often the available information is limited. In the "Guilty or innocent?" case, for example, one can argue that having a beard and a moustache are somehow associated and thus non-independent. However, an *ad hoc* conditional probability cannot be derived from the given information. On the other hand, in the "Should we install a thermoelectric power plant?" case, the yearly frequencies of events A and B are given in a double-entry table, making it possible to account for their association. However, that is not the case when independence over time is put to question, nor when the independence between coal quality and air quality is doubtful, as Bernardo pinpoints during the interview.

The problem is that independence is often a necessary assumption for probabilistic modelling (Borovcnik & Kapadia, 2011; Borovcnik, 2017). That is the case in the "Guilty or innocent?" learning environment. Computing the P-value in the juice experiment, the binomial model for the null hypothesis requires the assumption of independence between trials. In "Should we install a thermoelectric power plant?", independence is necessary to compute the probability of an environmental disaster in the long term. The assumption of independence takes part in inferential statistics as well, as one of the hypotheses of underlying probability theorems that formalise inference. That is the case of the Central Limit Theorem, which informs the construction of a 95% confidence interval in the "How many people attended the students' march?" learning environment (Chapter 8). In practice, the assumption of independence informs the sampling method.

Since, in general, $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$, the assumption of independence may have, as a consequence, the shrinking of probabilities that one might or not desire to be small, if

$\mathbb{P}(A|B) \geq \mathbb{P}(A)$. In Chapter 5, a prosecutor would like to depict the likelihood that an innocent person to match all features given by circumstantial evidence to be as small as possible. By arbitrarily adding features and multiplying their individual probabilities as if they were of independent events, $\mathbb{P}(\text{evidence}|\text{innocence})$ becomes reduced in favour of his case (Gigerenzer, 2002; Rosenthal, 2015). In Chapter 7, an environmental disaster is said to occur on a single day if four events happen at the same time. The original computation of the company assumes all of them to be independent and hence the probability of an environmental disaster follows from multiplying all four individual probabilities. In both cases, the shrinking of the probability at stake is of one specific party's interest, namely a prosecutor and an energy company. Respectively, a small probability can shape our perception of a suspect's innocence and a power plant's safety. It affects our judgement of probability, which is one of the fundamental issues in confronting risks (Levinson et al., 2012; Radakovic, 2015a, 2015b), whether it is the risk of a wrongful conviction or an environmental disaster.

Independence is a doubtful but necessary condition for making uncertainty and risks computable and, in the figurative contexts of the learning environments, probability calculations have consequences.

9.1.2 A probability-based epistemology

The use of P-values combines the use of conditional probabilities and probabilistic evidence, two of the building blocks of probabilistic literacy as conceived by Borovcnik (2017). Within the scientific paradigm, a P-value $= \mathbb{P}(D|H_0)$ provides evidence against a null hypothesis H_0 if the probability of some statistic D derived from collected data is small enough to doubt its compatibility with H_0. Many scholars (e.g. Gigerenzer, 2002; Borovcnik, 2006; White & Gorard, 2017) have pointed out the limitations of such type of evidence. For once, its correct interpretation relies on an FQT perspective, envisioning a repeated experiment that is often not realised. Moreover, there is a common misinterpretation of the P-value as if it were a more useful one, namely $\mathbb{P}(H_0|D)$, i.e. the probability for the null hypothesis to be true, given the evidence. The problem is not only that conditional probabilities are asymmetrical. If one is to obtain one conditional probability from the other via the Bayes' formula, necessary information of unconditional probabilities $\mathbb{P}(HD)$ and $\mathbb{P}(H_0)$ is often missing.

In Chapter 5, students have the opportunity to explore the construction of this type of probabilistic evidence in the juice experiment, and to discuss its application to a real judicial case, as a way to mathematise the principle of "beyond all reasonable doubt" (Greer & Mukhopadhyay, 2005; Rosenthal, 2015). The reflective potential in the learning environment lies, firstly, in the judgement of the appropriateness of transferring the reasoning from juice to justice, i.e. from an experimental study to an observational one (Samuels et al., 2016). Second, probability computations are based on a frequentist view on probability. Instead, their application in context is interpreted as an objective degree of belief, relying on the logical meaning of probability (Batanero & Díaz, 2007a; Batanero et al., 2016). Above all, the opportunities for critical reflections reside in how probabilistic models play the role of conveying evidence. In his historical and philosophical account, Hacking (1975) observes this duality throughout the whole modern history of probability. Probability is both related to stochastic laws of chance processes and to assess reasonable degrees of belief. Through its applications, probability can shape our opinion on a person's ability to tell sugary juice from others, or whether a person committed a crime or not. Probability affects our claims of knowledge.

9.1.3 Political arithmetic: Holding things together

The political aims of statistics can be traced back to its coinage as the science of the State (see Section 3.1). Donnelly (1998) argues that what made political arithmetic political was not only its aims but the conceptual framework of the State as a political unit. This category holds things together (Desrosières, 1991). This idea of aggregation into compartments of people takes place in several forms in the "How are PSU scores computed?" learning environment.

Students approach the notion of quantiles by relating it to income quintiles, to the first of which they aware of belonging. The admission system is designed in such a way that approximately 30% of the cohort cannot apply for higher education programmes. This fabricated normality provides tautological news of exclusion every year. PSU results are reported categorising students according to the type of school (public, subsidised or private), a variable that is strongly correlated with the socioeconomic class. In consequence, aiming to address achievement gaps produces double gestures of inclusion and exclusion (Andrade-Molina, 2017; Popkewitz & Lindblad, 2018). The use of graphical representations of statistical information about the State, in turn, divides national trends into partisan political groups.

In his problematisation of critical statistical literacy, Weiland (2017) points to the operationalisation of variables as one of the social dialectical negotiations at stake when producing and reporting statistical information. The work of Marilyn Frankenstein (e.g. Frankenstein, 1994) provides several examples of the problematic use of socially relevant numerical information, which "can furthermore create social categories and produce discourses that disadvantage individuals and groups" (Weiland, 2017, p. 43). The learning environment in Chapter 6 is undoubtedly providing a rich ground for this problematisation, by illustrating ways in which statistical approaches create categories of people moulded by distributions, the lens through which statisticians see variability (Wild, 2006). Statistics, therefore, can affect our sense-making of the inherent variability of the social world.

9.1.4 Probability and statistics in action

Perhaps the most overarching theme along the learning environments is that the mathematisation of variation, uncertainty and risks through probability and statistics and inform decision-making. As I discussed in Section 3.1, the driving force to the development of probability involved problems of decision making, starting with decisions of gambling in chance games. As extensive collections of numbers accumulated, the law-like patterns associated with gambling began to be transferred to socially relevant issues adjusted to statistical models. In our modern technological society, a mathematical grasp on the unknown and variant is not merely a way to describe or predict aspects of the world, but to prescribe and take courses of action, whether by automated technologies or by steering human will (Davis & Hersh, 1986).

In "Guilty or innocent?", probabilities are computed as an epistemological tool, to assess the innocence of a subject based on circumstantial evidence. As a consequence, probabilities are used to decide whether a person goes to prison or not. In "How are PSU scores computed?", we see the role of statistical procedures as an input to a sophisticated algorithm that selects and allocates high school graduates into higher education programmes. Moreover, reports on PSU results act as a political tool to herd public opinion and convey a compelling message that the political course must change. In "Should we install a thermoelectric power plant?", students experience the attempt of mathematising the risk of an environmental disaster, after a familiar experience with one that, at the time, could not be explained nor measured, except for its symptomatic consequences. Risks have many different meanings (Hansson, 2009), but, in principle, they imply coordination between the assessment of likelihoods and possible impacts (Levinson et al., 2012). The mathematisation of risk scenarios, rooted in probabilistic calculations, can make the difference in the quality of life of a community that wants to avoid living in an industrial park. In "How many people attended the students' march?", students can critically reflect upon the use of statistical inference to assess the reach of their exercise of the democratic right to gather and demonstrate. Their investigative work accounts for the sources of limited information, the assumptions of the full extent, and the uncertainty embedded in the estimation (Makar & Rubin, 2009). Social actors have different interests in validating or invalidating the cause of the demonstrations, and numbers can position them either as a fringe or a mainstream group.

All four learning environments are embedded in the reality that "no public decision, no risk analysis, no environmental impact, no military strategy can be conducted without decision theory couched in terms of probabilities" (Hacking, 1990, p. 4). Probability and statistics lead to action.

9.1.5 Emerging topics

Other topics emerged during the classroom experiences and further analysis. These do not relate directly to the concern of mathematics in action but could open up new approaches to investigate concerning the role of mathematics, probability and statistics in society.

Foregrounds

The notion of foregrounds is a preoccupation of critical mathematics education as well (e.g. Skovsmose & Valero, 2002; Skovsmose, 2011), and it is a way to acknowledge the political roots of learning obstacles (Skovsmose, 2005a). In Chapter 6, I have discussed the notion of foregrounds in light of the use of standardised academic tests and statistical tools. In the "How are PSU scores computed?" learning environment, the scene is set by a news article reporting ranges of scores according to types of school (public, subsidised and private), which, in the Chilean context, are highly correlated to socioeconomic groups (Valenzuela et al., 2014; Andrade-Molina, 2017). Students recognise a relevant aspect for computing PSU scores, namely quantiles since they identify themselves as belonging to the first income-quintile in order to apply for economic student aid, producing double gestures of (in)exclusions by making kinds of people (Popkewitz & Lindblad, 2018). However, at face value, students ground their academic achievements (background) and prospects for self-development (foreground) solely on individual preferences and abilities, illustrating how foregrounds can only be understood as part of the dialectic between structure and individual (Wedege, 2016).

Inter-viewing foregrounds require delicate dialectical approaches (Alrø et al., 2009). If they were to take part in the classroom as a preoccupation of critical mathematics education, the experience reported in Chapter 6 suggests that a statistics learning environment with a critical perspective can be appropriate to do so. However, an ethical dilemma lies in the eventual perpetuation of fatalism that characterises low-income students in Chile (Cavieres, 2011). As a possible and unwanted consequence, illustrating the systemic disadvantages at play in contextualised statistical activities, one may impact students' foregrounds negatively.

Data are constructed

Joel Best suggests that the popularised statistic-or-lie dilemma should be abandoned towards a construction-and-calculation bundle Best (2005), i.e. to realise that even the most honest statistics are social constructs (Berger & Luckmann, 1966) and scientific knowledge is the result of people's work. The classroom experiences highlight the relevance of not only reflecting upon statistical information (from data to claims) but of making one's own constructions.

In "Guilty or innocent?", students not only collect data for their juice experiment but first they construct it under criteria of reliability and validity. As a teacher and researcher, the experimental design allows us to build on the binomial model. A similar experience occurs with the "How many people attended the students' march?" learning environment, wherein an aerial photograph has to be treated in order to construct a source of data, before the actual collection in the form of random samples is carried out. Moreover, their context knowledge enables them to question the photographs as sources of data, since the dynamics and timing of the demonstration cannot be represented in an early-taken picture. Students experience the construction of sources of data.

In "How are PSU scores computed?", students first identify issues upon reading graphical representations of data, associated with mathematical, technological and critical reflections. When constructing their versions of a graph, the issues at stake are not only reconsidered but refined. Students experience that graphical representations are socially constructed.

In "Should we install a thermoelectric power plant?" the situation is different. Data are given in the learning environment, and students face no opportunities to check or revise it since it is also fictitious. Throughout the development of the sessions, students do question eventual "problems with data", but they refer to the use of those data in probability miscalculations, not the sources nor construction. Students do not experience the construction nor the collection of data.

The conception of data being constructed is relevant and informative of the critical justification, and it takes part in prominent frameworks of statistical thinking and literacy (Wild & Pfannkuch, 1999; Gal, 2002). However, it cannot be said to be specific about probability or statistics but to applications of mathematics in general.

Decision makers and stakeholders

Decision making is an abstract notion and a way of making it more concrete is to ask about who make decisions and who are the stakeholders of the consequences. Depending on whom are assuming the possible consequences of risks, they can be personal, institutional or societal (Borovcnik, 2015). However, this classification is not connected one-to-one with the respective decision-makers. Furthermore, when risks are of the societal type, whomever the decision-makers are, their legitimacy is at stake (Beck, 2000). Considerations of impact in risk situations can play a significant role in the discussion of stakeholders and decision-makers.

In the "Guilty or innocent?" learning environment, the principal stakeholder is the suspect, but a jury makes the decision. Students argue between the risks involved in a trial, in what can be identified as a discussion between type-I and type-II errors, technically speaking. On the one hand, the suspect can be declared guilty while being innocent. On the other hand, the suspect can be ruled as innocent when, in reality, he committed the crime. For the accused, a wrongful conviction has an immeasurable impact to be avoided at all costs. The jury faces no consequences.

In the "Should we install a thermoelectric power plant?" the possible consequence of an environmental disaster defines a societal risk, at least for the local community. In reality, the decision process involves the company and environmental authorities, whose headquarters are located in Santiago. In Chapter 7, I have pointed out many cases of students arguing against the power plant for the sake of the local community to which they belong. One student attempts to understand the case of the company:

CRISTIAN: Obviously, the purpose is to profit. The people who set the power plant, that is why they put it. Maybe it is not so much for the sake of the community.

Cristian's comment shows that, if the task is contextualised, the notion of decision making can be humanised. Decisions are not made in an abstract sense. There are social actors with different levels of power and different interests.

Though it is contextualised as a dispute over two parties' credibility, the "How many people attended the students' march?" inquiry is not an inquiry posed as a decision to be made. Nonetheless, it does illustrate the point that different stakeholders approach the problem differently, according to their interests. Students find their way to question the authorities' estimates from the sources of data to the non-transparency of procedures. However, one student makes the point that estimates are approximations and thus expected to be different. Moreover, using a statistical inferential approach based on random sampling offers yet another explanation for the variety of results.

The focus on decision-makers and stakeholders resonates with another preoccupation of critical mathematics education, namely the connection between mathematics education and ideals of democracy. The inquiry approach of the learning environments has its roots on John Dewey's (e.g. Dewey, 1916/1966) ideas scientific inquiry as a pragmatic democratic search for truth, by which an "education which organises itself in line with an inquiry process becomes 'education for democracy' " (Skovsmose, 1998, p. 196). However, Skovsmose (1998) makes the distinction between representative and deliberative, wherein only in the latter citizens take part in social and political decision-making. This distinction becomes evident in the project "Our community" (Skovsmose, 1994), whereby students not only make budget decisions in the classroom setting, but they get a visit from their Mayor and express their points of view. Valero (1999) reports on a group of Colombian mathematics teachers participating in deliberative interaction to support a national educational reform. These and other cases show how the ideals of democracy do not rest on the classroom imagination of agency, but that a further step can be to engage in actual deliberation.

None of the four learning environments gives opportunities to engage in out-of-school deliberation, but some of them could. In "How are PSU scores computed?", students not only reflect upon a graphical representation with political aims, but they construct their versions and reinterpret the claims behind PSU score trends, which could be shared in some media outlet as a response. This response is precisely what an expert did (Fábrega, 2017) and, as a consequence, the mentioned campaign to "save the emblematic [schools]" began to lose momentum. Students have the tools to engage in that national debate as well. In "How many people attended the students' march?", students make different estimates of the crowd size under different assumptions which are transparent and could be shared in some media outlet. As I commented in Chapter 8, the women's march in Santiago set a precedent in responding to authority, where experts provided their estimates of the crowd as being far more massive than the one given by the police. Students could eventually do the same. The other learning environments are more challenging to extrapolate outside the realm of the classroom. In "Guilty or innocent?", it is hard to imagine a way of engaging in deliberation in the Chilean context, where there is no jury making decisions. In "Should we install a thermoelectric power plant?", in principle, activism is possible and actual. However, the fact that only the event is authentic limits the students' possibilities to have a credible voice in the real struggle.

Perhaps, it is too much to ask students to change the world in a classroom experience. Nevertheless, decision-making in teaching practice can help them question their active role as future citizens. That is, for example, the case of a student in a study of teaching risk associated with nuclear energy (Radakovic, 2015b, p. 283):

TALIA: But if there is pollution doesn't it have to be addressed by the government?

HELEN: People can go against the government.

Helen may not be confronting the authority, but she just realised that doing so is a possibility.

9.2 Reflective knowing and frameworks for analysis

A critical justification for teaching and learning underlies the claim that an educational practice can, in fact, enable students to live up to the critical competence (Blum & Niss, 1991; Niss, 1996). I have taken an overarching interpretation of this competence from Skovsmose's entry points on reflective knowing (Skovsmose, 1992, 1994) as a way to separate mathematical, technological and critical reflections may be. Overall, critical reflections are elicited by students discussions (Barbosa, 2006), if they refer to the way applications of probability and statistics shape our conceptions and actions upon some aspect of the world.

I have already mentioned some of the themes of the formatting power of probability and statistics that the learning environments have the *potential* to take part in critical reflections. Some other topics have emerged, not exclusively related to probability or statistics. Let us review to what extent the four learning environments show evidence that students *did* or *could* reflect critically, based on the different specific analytic approaches taken.

9.2.1 Probabilistic evidence in justice

The “Guilty or innocent?” learning environment was meant to provoke critical reflections on the use of probabilistic elements as evidence in real trial cases (Greer & Mukhopadhyay, 2005). The case examined is the one of Malcolm Collins (People v. Collins, 1968), examined by Rosenthal (2015) around three issues: multiplication of probabilities, the prosecutor’s fallacy (Thompson & Schumann, 1987) and the out-of-how-many principle. The learning environment is designed to reflect upon this case in the third session.

As reported in Chapter 5, students to engage in reflections connected to the three issues pinpointed by Rosenthal (2015), but there is no clear evidence of them to be of the critical kind. The multiplication of probabilities is addressed via mathematical reflections in a right-wrong divide (Skovsmose, 1994). The out-of-how-many principle takes the form of a technological reflection, addressing the appropriateness of using a probability of $1/12,000,000$ in a context of $6,500,000$ people.

Opportunities for critical reflections reside in the use of the probability $\mathbb{P}(\text{evidence}|\text{innocence})$ as convincing for a jury to declare the suspect guilty beyond a reasonable doubt. The prosecutor’s fallacy plays a role if the probability is misinterpreted as $\mathbb{P}(\text{innocence}|\text{evidence})$, upon which students do not reflect. Other reflections approximate the critical stance towards probabilistic evidence. Pablo advocates against its use “because there is still a possibility that he is innocent. There are still more people with those characteristics”. Another discussion takes place about the type-I and type-II errors, as in whether it is worse to convict an innocent person or to acquit a criminal. However, none of these discussions reaches the scope of the reflecting upon the formatting power of probability; students do not acknowledge how probability creates reality beyond the case before them.

After the experience with the following three classroom experiences, the lack of evidence for critical reflections can be explained, in part, by the assignation of time to different activities. That is not to say simply that more time would have led to critical reflections. Two out of three sessions had the juice experiment as the guiding investigation. The case of Malcolm Collins is approached by the students passively, as consumers and not producers of data-based arguments (Weiland, 2017). They do not collect necessary data, nor compute probabilities themselves, nor make a decision on the legal case. The other three learning environments show that the popular principle of “learning by doing”, a virtuous connection between education and democracy through inquiries (Skovsmose, 1998), can be experienced.

Another crucial factor in the inability to critically reflect upon the case, in the sense of Skovsmose, is that the three elements of the local framework for analysis are approached in a right-wrong manner. The reason why Rosenthal (2015) presents the case of Malcolm Collins and other legal cases to make his point, is that the “when to multiply” question, the prosecutor’s fallacy and the “out of how many” principle are wrongfully applied. The example, therefore, invites to examine what is wrong, correct it, and not to critique the fact that mathematics is playing a role.

9.2.2 Statistical messages about standardised educational assessment

“How are PSU scores computed?” is a rich learning environment, from which a variety of critical topics could emerge. These include the double gestures produced by statistical information in students foregrounds (Popkewitz & Lindblad, 2018), the use of the normal distribution to shape social matters, and the political framing of graphical representations in the media.

In order to address the guiding investigation in the learning environment, students need to have a grasp on position measures to make use of the percentiles in Table 6.2. Students make a connection to income quintiles in the population, as they are aware of belonging to the lowest, which enables them to apply for economic aid for higher education. To some extent, Claudia reflects on the prescriptive function of mathematical models (Davis & Hersh, 1986) in society. She acknowledges that, for example, people in the fifth quintile are not eligible for grants or loans "because there lies society's highest rank, i.e. they have more money, i.e. they have the means to pay for their studies." This episode can be considered to be a critical reflection since Claudia is becoming aware of how this specific statistical tool sorts, splits into arbitrary groups the cohorts of higher education applicants and assigns resources accordingly.

Further prompting could engage in a discussion at a more systemic level, based on David's rationale about his foreground. Higher education is seen as a good one should pay for, whether from one's pocket or by obtaining a student loan, and free healthcare is a privilege to the armed forces. Similarly, when Gal (2005) introduces a notion of probabilistic literacy, he exemplifies it with a discussion on the possibility of acquiring health insurance, without suggesting to ask whether healthcare could be a guaranteed right. Systematic structures shaped by probability and statistics may be a further step in the development of literacy notions, as Weiland (2017) introduces in his critical literacy framework.

By openly analysing graphical representations of statistical information in the media (Figure 6.4), researcher, teacher and students develop together a proto-framework of aspects to be aware of when reading graphs (Table 6.4). Two of these aspects can be considered evidence for critical reflections, involving a graph on the newspaper portraying trends on PSU scores in two specific schools (Figure 6.3).

The first aspect corresponds to an awareness of how the graphical representation shapes our perception of the data represented (G5), beyond the correct extraction of information at face value. Néstor reacts by claiming that "if people saw it like this, without the numbers, they would say: 'it was a lot' (...) at least when I first saw it, I said 'Wow, it decreased a lot', but then I saw the numbers and (...) When one sees things without analysing, one sees... one sees a lot". His reflection is in line with the fifth entry point on reflective knowing (Skovsmose, 1994). Néstor is becoming aware of the effects of depicted statistical information on his and other people's conception of a particular phenomenon.

The other aspect I interpret as critical is the production of alternative interpretations of the same graph (G6). Beyond reflecting upon how the representation can shape our perception of the school system, Miguel challenges it by proposing that the trend could even mean good news "because maybe it is not a problem with those schools. Perhaps they stay the same, but there are other people doing better". Néstor is more specific on the grounds of this reflection, by explaining that "if they decrease, they lower their percentile and others go up". These exemplary critical reflections are rooted in understanding how PSU scores are computed and reported on a relative scale. As much as I have focused on the fifth entry point on reflective knowing (Skovsmose, 1992, 1994), this type of reflection is more consistent with Skovsmose's version of mathematical literacy. For him, *mathematcy*, "tentatively, (...) can be understood as a kind of competence in which mathematics-oriented, model-oriented and context-oriented reflections are brought together as an epistemic unit" (Skovsmose, 1998, p. 200). Miguel and Néstor are engaging in reflective knowing as a whole.

One more element that stands out in this classroom experience is that the critical aspects of the proto-framework were only elicited once students produced their versions of the graphical representation of the same data featured in Figure 6.3. In general, statistical literacy is called for in reading contexts, i.e. assuming citizens to be consumers of statistical messages (e.g. Gal, 2002; Watson & Callingham, 2003). Weiland (2017) problematises the duality between consumer and producer, proposing that learning to read the world with statistics can enable them to write it. The experience with "How are PSU scores computed?" suggests that it can be the other way around. By exercising the production of statistical messages (e.g. a graphical representation), they become enabled to refine their critical consumption of statistical messages.

9.2.3 Probabilistic thinking and literacy in the risk society

For the "Should we install a thermoelectric power plant?" learning environment, I have used the categories of probabilistic thinking and literacy by Borovcnik (2017) to analyse students' reflections, connecting them to the sociological concepts of the risk society (Beck, 2000). As discussed at the end of Chapter 7, students engage in many different reflections of the mathematical and technological kind (Skovsmose, 1994; Barbosa, 2006). In retrospect, some reflections do have the potential of being critical, since they are context-oriented (Skovsmose, 1998), but do not refer to the use of the probabilistic tools to approach the inquiry.

Critical reflections could take place if students made explicit connections between the formal probabilistic notions and how their applications affect their part of the world. In the context of risk, based on the work of Levinson et al. (2012), Radakovic (2015a) proposes five elements to be considered in a pedagogy of risk: knowledge, beliefs and values; judgement of impact; judgement of probability; representations; and estimation of risk. Some of the students' reflections on "Should we install a thermoelectric power plant?" could be critical in connection to those elements.

When Cristian questioned the meaning of $\mathbb{P}(C)$ expressed as 30% poor-quality coal, his reflection is technological, in the sense of the appropriateness of one or other interpretation to solve the problem of addressing the risk of an environmental disaster. However, the discussion could follow up. It could address how this application may be influencing the judgement of the probability of a disaster. These reflections and others make use of a frequentist meaning of probability (Batanero & Díaz, 2007a), which, after some prompting, could lead to reflect on the appropriateness of such an interpretation upon one-off decisions (Borovcnik, 2015). In other words, students could realise that we are interpreting the *chance experiment* that happens every day, conflicting with the fact that the power plant is installed only once.

The first part of the task required students to correct the company's calculation of the probabilities of a set of events, which affected the resulting probability of a single-day disaster by the multiplication rule. That process gives an opportunity to critically reflect on how these approaches shape our judgement of probability or how does the company makes use of small probabilities to influence the evaluation of risk.

In light of the analysis, I cannot claim strong evidence of critical knowing taking place in students reflections during the "Should we install a thermoelectric power plant?" experience.

Part of it can be explained by the relation between the learning environment's structure and the scope of analysis. Based on the principles proposed by Blomhøj (2016) for inquiry-based learning, a joint reflection is planned at the end of the sequence during Session 3. Kuntze et al. (2017) also have recommended to let give a reflection-oriented instructional framing. This way, students have the opportunity to reflect on the task they have faced. For the sake of prioritising time resources, this is the one part of the experience that is fully coded for the analysis. I have reviewed and listened to some of the group conversations to make sense of their results and reflections at the end, but critical reflections may have occurred in the early process. Moreover, the joint reflection at the third session is guided by three questions, two of which are direct requests for results.

Reflections that can be considered much closer to being critical take place in the decision making part of the sequence. These are embedded in categories of probabilistic thinking of "separating reflecting and making a decision" and "balancing formal and psychological aspects" (Borovcnik, 2017). Pedro realises that "low joint probabilities, they produce a big disaster" and that "ideally there should be no probability for an environmental disaster to occur". Students take impact considerations, such as cost/benefit and the perception of risk as a hazard to be avoided at all costs. They include aspects of the risk society, such as going beyond formalities and ask themselves how do they want to live, which encapsulates the duality of risk statements as being factual and of value Beck (2000).

In my view, a reflection that stands for its critical character is Sandra's statement: "I mean, the fact that it (the probability) is small, it does not mean that nothing will ever occur here". It may not be explicit, but she is separating the mathematical approach of "low probability" with the statement "nothing will ever occur here", thus acknowledging the effect of low probabilities in the evaluation of risks as conveying a sense of security. In the analysis, her comment is coded as part of "the ability to balance between psychological and formal elements" (Borovcnik, 2017, p. 1497), since her argument is *anchored* (Tversky & Kahneman, 1974) to the recent disaster of Quintero and Puchuncaví, but taking the low probability into account.

In his article, (Borovcnik, 2017) pursues to make a comprehensive framing of elements of probabilistic thinking and literacy in the context of risk. In the analysis, the literacy aspects do not play a significant role in identifying instances of critical reflections, since the theoretical character of probability and independence, conditional probabilities and probabilistic evidence configure themes or topics of conversation. On the other hand, the elements of probabilistic thinking he describes are posed as "competing for intuitions and strategies" (Borovcnik, 2017, p. 1496), which resonate more with the character of the risk society as a society that reflects on itself (Beck et al., 1994). Critique does not necessarily mean to have a clear stand, nor a dogmatic set of values. On the contrary, a critical stance towards mathematics means realising it is not good, nor bad, nor neutral (Skovsmose & Valero, 2012). Critique implies embracing this uncertainty (Skovsmose, 2008), and the ability to acknowledge, separate and balance competing probabilistic intuitions in risk contexts goes in that direction.

9.2.4 Critical statistical literacy and crowd sizes

The framework for critical statistical literacy proposed by Weiland (2017) has a particular focus on acknowledging and separating what it means to read and write the world with statistics. I have pinpointed in the closing of Chapter 8 that the coding is generous, as the categories in the framework are very ambitious. Nonetheless, some of the episodes coded in the "How many people attended the students' march?" learning environment are worth highlighting as evidence for their *potential* to follow up into critical reflections.

First, the issue of moving the statistic-or-lie divide to a calculations-and-constructions bundle (Best, 2005) is more or less explicit in the discussion between Andrés and Gastón advocating that the media lie when reporting estimates for the number of attendees. At the same time, César, based on his work, challenges that interpretation by advocating for estimates to be own constructions and approximations. This episode is coded both as *making sense* and *reshaping* from Weiland's framework. It can be considered evidence for critical reflections, since, beyond the technological appropriateness or necessity for solving the inquiry, Gastón says that "they take advantage of the situation", thus becoming aware of the power of this application of mathematics in people's perceptions when reading numbers on the media. Moreover, César makes sense of statistical estimation based on his own experience constructing data and random sampling. Therefore, not only can students be aware of how they use mathematics shapes public opinion, they can see how the statistical idea of variation is involved.

Another relevant episode that stands out is the realisation by Silvia that the march is moving, which may lead to an overestimation if the extrapolation from the first close up picture is made considering the full path of the march. Aníbal and his group react by mentioning their construction of three different intervals based each on different assumptions, unprompted (see also Figure 8.9). They realise that there is a dialectical tension in the interpretation of the full extent of the demonstration, so they negotiate it. To a large extent, this discussion is evidence for technological reflections, since they are discussing the appropriateness of their assumptions to answer the crowd size inquiry. However, Aníbal follows up by saying that their middle ground estimate is between 40,000 and 61,000, "so we believe it is still pretty high" contrasting the 30,000 attendees which authorities had estimated. Beyond the calculations and negotiation of assumptions and procedures, Aníbal's comment displays an awareness of how different approaches shape his perception of how massive the demonstration was.

The third episode I would like to highlight is the one of casting shadows over sources of data. Students use their context knowledge to reject the pictures as valid sources of data since they are taken too early to be representative of the march's attendance. These aspects include the presence of buses, the absence of police and the direction of shadows. Once again, these reflections are technological, in the sense of evaluating the appropriateness of the data sources to estimate the crowd size. However, this episode is followed by Gastón yelling spontaneously "the institution lies!", leading to the first episode mentioned above. Therefore, it is undoubtedly an input that can provoke critical reflections.

Two of these selected episodes have something else in common concerning the analysis. They are coded under two categories, of reading and writing the world. The debate between lies and estimates is both *making sense* (reading) and an effort to reshape (writing). Casting shadows over data sources is both *evaluating* the sources (reading) and *communicating subjectivity* (writing). Students are making sense and evaluating anchored in their statistical investigations. The framework proposed by Weiland is based on a problematisation of the usual focus on reading contexts, portraying citizens as readers of statistical messages (e.g. Gal, 2002). He is inspired by the ideas of Gutstein of "using mathematics to change the world" (Gutstein, 2006, p. 27) to include the categories of writing the world with statistics, since "reading the world with mathematics can, in turn, lead to writing the world with mathematics" (Weiland, 2017, p. 38). Using this framework to interpret students reflections, I suggest that the reciprocal is true in a pedagogical sense. Exercising an inquiry approach, writing the world with statistics can, in turn, lead to reading the world with statistics as well, thus being consistent with the ideas of many scholars on the virtuous connection between inquiry learning and ideals of democracy (Dewey, 1916/1966; Artigue & Blomhøj, 2013; Skovsmose, 1998).

9.3 Tensions between the design principles

A central element of the *problématique* concerns the design of learning environments that resonate or enable the critical justification. As mentioned in Chapter 4, the principles of *exemplarity*, *inquiry approach* and *pragmatism* not only informed the designs expressed in the learning environments, but represent dimensions to focus on for a retrospective evaluation. In this section, I will address some of the tensions encountered between these principles during the classroom experiences.

9.3.1 Exemplarity and inquiry approach

The "Guilty or innocent?" learning environment is an interesting opportunity to explore the tension between *critical exemplarity* and the *authenticity* embedded the *inquiry approach*, since two contexts are involved in one: the juice experiment and the case of Malcolm Collins. Both cases are exemplary from an *instrumental* perspective, by developing notions of probabilistic evidence in the form of – intuitively approached – P-values. However, the *critical* and *subjective* perspectives are compromised.

On the one hand, students design the juice experiment, collect data and discuss their results, turning the investigation into their personal experience (Dewey, 1938/2015). The experiment does not account for the use of probabilistic evidence in socially or politically relevant issues, namely the justice system; hence it is not critically exemplary. On the other hand, the case of Malcolm Collins is carefully selected for being an icon of the questionable uses of probabilistic procedures in justice (Greer & Mukhopadhyay, 2005; Rosenthal, 2015). My impression is, however, that it is difficult for students to engage with the case at a personal level. The case occurred far from their reality, both geographically (United States) and historically (the 1960s).

Moreover, Chile's judicial system does not work with juries to be persuaded. As discussed at the end of Chapter 5, critical reflections are not engaged in the classroom experience, despite the critical exemplarity of the real case. In their commentary on the humanistic dimensions of probability, Sriraman and Lee (2014, p. 119) warn about the resistance on behalf of policymakers on the use data "perceived as being too real (and grim)". In the judicial case, the resistance to authenticity comes from the students. However, my position is that the problem lies not in the authenticity of the evoked event or the availability of data; it is the ability of the design and implementation to make the case exemplary.

The guiding inquiry of "Should we install a thermoelectric power plant?" comprises an authentic event and question, as opposed to data and aspects of the solution. As pinpointed in Chapter 7, the fictitious elements do not limit the potential to provoke critical reflections, but the scope could be broader if students could collect real data and explore the authentic complexity of an environmental evaluation. To be more precise, the learning environment does allow students to experience its *critical exemplarity* regarding concepts of the risk society (Beck, 1992, 2000), but not about the construction of data (Best, 2005). The paradox is that a critical justification calls for students to reflect upon the analysis and evaluation of real exemplary cases of applications of mathematics. Reflections take place, but the case is made-up and likely oversimplified.

As pinpointed in Chapter 8, an interesting dilemma of authenticity takes place in the experience with "How many people attended the students' march?". The approach towards the inquiry during Session 2 means to be *instrumentally exemplary* for the big statistical idea of sampling (Garfield & Ben-Zvi, 2008; Ben-Zvi, Gravemeijer, & Ainley, 2018) and, in particular, of the effect of sample size on the sampling distributions. The authentically unpredictable character of collecting own data resulted in not making a strong case for this effect. Though some researchers call for tailoring data sets to fit the desired results amid variations (e.g. P. Cobb et al., 2003), a balance can be found by working in collaboration in the classroom to produce large numbers of data that favour convergence. If outliers or unplanned-for variations appear, they should be part of the discussion. After all, authentic contexts and different sources of variation are at the core of statistics (G. W. Cobb & Moore, 1997; Moore & Cobb, 2000).

9.3.2 Inquiry approach and pragmatism

The main tension between the inquiry approach and pragmatism is experienced in the interventionist role of teachers and researcher to support students' work. In an ideal case, we wanted to develop a practice whereby students discovered the curricular content to be covered while engaging in the inquiry, given the limited time frame. During the interview with Carla, one of her comments illustrates this tension very clearly:

> CARLA: (I mean... I like structure very much because it allows any person to follow the class. Now, that structure needs not to be so rigid, and one can be making changes because there will evidently emerge excellent ideas, different from what one had prepared. But something too ambiguous, too broad, with no directives, may have them (students) reach no answer.

A relief to this tension is the dialogical feature of *identifying* (Alrø & Skovsmose, 2002), by which teachers may help to *crystallise mathematical ideas* in order to advance the inquiry. For example, in "Guilty or innocent?", a more expository session was needed to compute the probability $\mathbb{P}(X \geq 8)$, necessary to build probabilistic evidence applying the binomial model, scaffolding from the case of two and three trials. Similarly, in "How are PSU scores computed?", the second session was oriented to compute PSU scores under three different cohort scenarios. At the beginning of the second session of the "Should we install a thermoelectric power plant?", students share their takes on the calculation of $\mathbb{P}(ED)$ to reach a consensus, facilitated by explicitly asking them to make explicit their judgements about the independence of events. We fail to do the same with the probability of an environmental disaster in the long term. While I do challenge students to reconsider their proportional approach once it leads to a probability of 2.7, we do not intervene further. In "How many people attended the students' march?", in my opinion, Carla managed to navigate this tension very well for the construction of confidence intervals. She knew when and what to do with students' input, and when to move the discussion to the whiteboard.

In hindsight, the tension between these principles arises from two misunderstandings, on my behalf.

First, models of inquiry-learning provide certain principles which need not be interpreted as a strictly chronological structure. For example, I have drawn upon the model of scene-setting, students' independent work, and joint reflection (Blomhøj, 2016) to give an initial structure to the designs, wherein the systematisation of mathematical notions is only referred to in the latter. The need for flexibility is consistent with the use of the five steps model of Bybee et al. (2006) (engage, explore, explain, elaborate and evaluate) by (Radakovic, 2015a, p. 318), who emphasises that "while presented linearly, the five steps do not always proceed chronologically and each of them may contain the teacher's help and guidance".

On that same line, the second aspect to consider is that an inquiry approach is not the same as a minimal guidance approach. Hmelo-Silver et al. (2007) argue that inquiry learning should include constant scaffolding and assessment, which other more radically constructivist approaches do not (Kirschner et al., 2006). Though possible mathematical pitfalls can always be planned for, there is no general rule, and the criteria for real-time intervention is rooted in the relationship built between teacher and students:

CARLA: You have to know your class and say: "Oh, I think, at this point, I could pose this open question and let them keep investigating".

9.3.3 Pragmatism and exemplarity

To some extent, examples may not be necessarily approached through the intended curricular coverage, building tension between *exemplarity* and the curricular aspects of *pragmatism*. Some of the learning environments can illustrate this as they were carried out.

In "How are PSU scores computed?", we had agreed with Alejandra to avoid mentions to the normal distribution and approach the calculation of PSU scores through percentiles instead. This negotiation still allows the case to be exemplary from a subjective perspective, but the other perspectives of exemplarity are compromised. The notion and use of theoretical distributions are among the key statistical big ideas (Garfield & Ben-Zvi, 2008) which we did not exemplify from an instrumental perspective. From a critical perspective, the case could have been exemplary of the distinction between phenomena *being* normally distributed or *designed* to be (see Section 6.2). However, the use of empirical distributions for ordering and classifying people still holds. Similarly, in the "Should we install a thermoelectric power plant?" learning environment we face the challenge of aiming to use the binomial model for computing the probability for a disaster to occur in the long term. However, the specific probability calculation at stake is a particular case that only requires the application of the multiplication rule and the probability of the complement, as seen in Equations 7.3 and 7.4. Students do not reach those developments either, but still find their way to discuss, for example, the assumption of having the same disaster probability each day. In these two classroom experiences, the balance is tilted against the pragmatic principle of curricular coverage.

In "Guilty or innocent?" and "How many people attended the students' march?", the tension is approached differently, by pausing the inquiry and crystallise mathematical ideas, namely the computation of binomial probabilities and the construction of normal confidence intervals, respectively. Here we became closer to the upper bound of curricular coverage under the pragmatic principle, but they are not necessary to advance the critical exemplarity of the cases. The case of Malcolm Collins (People v. Collins, 1968; Rosenthal, 2015) can be approached applying the multiplication of probabilities rule with no mention of the binomial model. The latter was necessary for the juice experiment to grasp the production of probabilistic evidence, thus being instrumentally exemplary. The estimation of crowd sizes via random sampling and confidence intervals is not necessary. Students proposed several approaches that are closer to what some experts do, i.e. to define estimates for people density according to how much one can move in different areas.

A balanced relaxation of a pragmatic adherence to the curriculum could be achieved by echoing the call for statistics education researchers to focus on big statistical ideas instead of specific procedures (P. Cobb & McClain, 2004; Ben-Zvi, Gravemeijer, & Ainley, 2018). A compatible ongoing research agenda in the field is that of informal inference, which opens up the notion of inference as a type of reasoning that encompasses the use of data, an explicit account for uncertainty and claims of generalisation (Makar & Rubin, 2009). Some of the inquiries leading the learning environments in this thesis are compatible with the kinds of informal inference problems suggested by Zieffler et al. (2008). For example, "Guilty or innocence?" corresponds to a comparison between two models (art versus chance), whereas "How many people attended the students' march?" is precisely made as an inference about the population from a sample. Though authors recognise that informal inference is a working definition that supports research, it can guide the teaching design by focusing on the three dimensions proposed by Makar and Rubin (2009). Be that as it may, outside a research project, the flexibility to compromise mathematical content to highlight other qualities of an example is on the hands of each teacher.

In this chapter, I have revisited the four classroom experiences from the three relevant perspectives of the general *problématique*. Regarding the preoccupations of critical mathematics education, I have revised how the examples illustrate the formatting power of probability and statistics and other emerging topics. Concerning the validation of the critical justification, I have reexamined the analyses and how they provide evidence for students' critical reflections. As for the teaching practice, I have revisited the main tensions encountered between the design principles.

The ground is paved for addressing the research questions and general *problématique* of coherence in the concluding remarks that follow.

Chapter 10

Concluding remarks

Throughout this thesis, I have explored different aspects of the critical perspective on teaching and learning probability and statistics in upper secondary school. The empirical basis of the thesis is anchored in Chile, analysing texts from the curriculum and the design, implementation and analysis of four learning environments, in collaboration with three mathematics teachers.

In this chapter, I give closure to the project. First, I avoid suspense and address the five research questions. Second, I draw on the research questions to travel along the general *problématique* of coherence between the critical justification, possibility and implementation.

10.1 Revisiting the research questions

My approach in the first study was to let curricular documents speak for themselves, in the sense of not looking for *a priori* interpretations of how notions of decision making take part in the positioning of social actors and legitimation for the teaching and learning of probability and statistics. For this reason, despite raming the thesis within the *problématique* of Critical Mathematics Education, the research questions belonging to this part of the research project have a loose formulation in terms of what functions or meanings I seek. That is the case of **RQ1**:

RQ1 What is the role given to notions of decision making in the Chilean and Danish upper secondary school probability and statistics curricula?

In light of the analysis, there are a few key findings. Forgiving the redundancy, decisions are made by decision-makers. Most of the texts position students in such a role, but not in the present. Instead, it is expected that education will enable them to make decisions eventually. In Chilean texts from the current curriculum, (MINEDUC, 2015b, 2015a), teachers' actions are the ones enabling students to make decisions, and it shifts to the new curricular proposal (MINEDUC, 2017), where making decisions is itself a learning goal.

Decisions are embedded in discourses of legitimacy (van Leeuwen, 2007) for the teaching and learning of probability and statistics in many ways. For once, there is a narrative of rationality given to decisions informed by work in probability and statistics, evidenced by collocations such as "grounded decisions", "decisions with evidence" and "decisions with information".

Legitimacy is also conveyed by appealing to authority. There are mentions of statistical data-based political decisions to denote the relevance of learning. Notions of probabilistic thinking and statistical literacy are not unfolded, but represent intertextuality with educational research communities.

The comparative interest reported in Section 3.3 is due to a real and a good reason. The real one is the suggestion by reviewers to make my research of interest once published in a Nordic journal (Elicer, 2020). The good one is that it allows me to open my perspective and describe the Chilean case compared to the Danish one. The logic of comparison in research "implies that we can understand social phenomena better when they are compared in relation to two or more meaningfully contrasting cases or situations" (Bryman, 2015, pp. 64–65). Thus, the second research question is formulated as:

RQ2 Which similarities and differences can be drawn from these two cases?

There are two notable differences between the cases. First, Danish texts position students as future citizens that make decisions in a broader scope of contexts, namely social, everyday, professional and study context. Whereas, in the Chilean case, the contexts of decisions are omitted or abstracted. However, looking at a broader picture, the upcoming curricular bases recognise, from public diagnostic surveys, that students "demand protagonism in decision making and aspire to contribute to solving problems in the world they live in, such as poverty eradication, climate change and sustainable development" (MINEDUC, 2017, p. 13). Second, regarding legitimation strategies, though both curricula draw on the work of professionals as an appeal to authority, only the Danish case makes it explicit. My interpretation of authorisation in the Chilean texts relies on the use of terms such as "probabilistic thinking" and "statistical literacy". In the Danish case, one of the analysed official sources is authored by three experts (Ekstrøm et al., 2017).

The main finding is that, in both curricula, part of the purpose of mathematics education in general, and probability and statistics in particular, is to enable students to participate in decisions of social and political relevance. However, it does not seem that the connection between decision making and the subject matter goes beyond the personal psychological scope.

One of the limitations of the approach taken and reported in Elicer (2019, 2020) is the elusion of one of the dimensions of Fairclough's version of Critical Discourse Analysis. The three-dimensional model portrays texts as factual examples of discourses, which, in turn, are the mediators of broader social structures. Fairclough (2010) argues that CDA can be seen as a methodology for Critical Research. It is a variant of Bhaskar's explanatory critique (Bhaskar, 1986), formulated in four stages: "(1) focus upon a social wrong, in its semiotic aspect; (2) identify obstacles to addressing the social wrong; (3) consider whether the social order "needs" the social wrong; and (4) identify possible ways past the obstacles" (Fairclough, 2010, p. 235). Fairclough is quite insistent in avoiding a mechanical interpretation these stages, since, for example, It is also helpful to distinguish the 'theoretical' and 'procedural' from the 'presentational' order one chooses to follow in, for instance, writing a paper" (p. 234). In that sense, the bottom-up approach I took – from texts to discourses – is not at all dissonant with the methodology. However, the primary goal of CDA is to understand broader social systems, forms or orders that configure social "wrongs" (e.g. poverty, racism, gender discrimination). In retrospect, can I say anything about broader social structures or social wrongs?

If I dare to look at the broader picture, the emerging discourses echo one a trend characterised by Valero (2017) as technological optimism, in a two-fold manner. On the one hand, texts display praise on the role of rational and data-based decisions on every scope, including the social and political, modelled using probabilistic and statistical constructs. This optimism is embedded in what Popkewitz (2004) calls the alchemy of mathematics; the belief that mathematics can turn anything into gold. On the other hand, texts are optimistic about the role of education as a technology. According to Biesta (2009), education has three overarching functions: qualification, socialisation and subjectification. The analysed texts relate to the latter, as they take part in discourses that encourage students to become particular kinds of persons, autonomous decision-makers whose actions are rooted in the clean rationalities of mathematics. In principle, these ideals and beliefs do not strike me as social wrongs. However, whichever the intentions put forward to advance access to mathematics “for all”, the declared importance of mathematics on the fabrication of modern citizens is a double-edged discourse that excludes those who do not participate.

Let us come back to the general *problématique*, starting with the justification problem (Niss, 1993, 1996). In hindsight, is the critical justification activated as a “critical competence” argument in the texts? Or, is it at least compatible? Let us come back to the formulation of such an argument:

> The *critical competence* argument focuses on preparing students to live and act with integrity as private and social citizens, possessing a critical competence in a society the shape and functioning of which are being increasingly influenced by the utilisation of mathematics through applications and modelling. Such a critical competence aims to enable students to *see and judge* independently, to recognise, understand, analyse and assess representative examples of actual uses of mathematics, including (suggested) solutions to socially significant problems.
>
> (Blum & Niss, 1991, p. 43)

From the formulation, one can deconstruct the argument into three claims. First, there are aspects of the social world that are shaped and function utilising applications of probability and statistics. Second, it is possible to bring examples of those applications into school learning environments. Third, students can critique and eventually suggest solutions to socially significant problems.

The first element in the argument is the assumption that aspects of society are shaped and function using mathematical models and applications. I have discussed in Chapter 3 that the development of probability and statistics has been intrinsically connected to problems of decision making (Hacking, 1975, 1990; Borovcnik & Kapadia, 2014; Mateos-Aparicio, 2002, 2004), and increasingly gaining territory from gambling decisions between players, to decisions in the application of the law, determination of scientific evidence, development of economic theories, and policymaking. In contrast to the Danish case where “political decisions” are mentioned explicitly, the Chilean texts acknowledge that there is a world that comprises “situations of uncertainty” and surrounded by “statistical information” in a broad sense.

The second element in the argument is that there should be representative examples that can be worked in the classroom. There is, at least, one mention to the importance of teachers promoting “contextualised learning” and the development of “probabilistic thinking” and "statistical literacy". The latter notion presumes a connection to real-life applications of statistics (Gal, 2002). Therefore, in principle, it can be assumed that the texts encourage the use of real-life examples of applications of probability and statistics in school activities.

Finally, there is the question of what those exemplary cases in classroom activities enable students to do. For once, the texts encourage students to critically "evaluate" and "analyse" statistical information in the media, a perspective that suggests a direct connection to the critical competence argument. Moreover, the texts encourage students to make decisions "under uncertainty", "grounded in statistically significant information", enabled by the development of "probabilistic thinking" and "statistical literacy". From a critical argumentative perspective, these decisions should refer to socially relevant issues. As I have stressed, the curricular texts suggest a connection to personally relevant decisions alone. Drawing upon the "insurance" example of Gal (2005), contextualised learning, the development of probabilistic thinking and statistical literacy would enable students to decide whether and when to buy health insurance. However, they are not invited to imagine or suggest solutions to the socially relevant problem of access to healthcare. While the new curriculum recognises the will for future citizens to have an incidence in matters of political, social and environmental decision making, this is part of a general appreciation in the curricular bases, and not connected to probability and statistics (MINEDUC, 2017, 2019). If any, these formulations sound more aligned with the *utility argument*, which emphasises that "mathematics instruction should prepare students to utilise mathematics for solving problems in or describing aspects of specific extra-mathematical areas and situations" (Blum & Niss, 1991, p. 43). The curriculum recognises the social scope of uncertainties and statistical information. However, including such scenarios in the classroom is encouraged to facilitate students to acquire helpful tools and make decisions for their own sake.

Concerning the justification part of the *problématique* this thesis is embedded in, it is safe to say that the (Chilean) curriculum does not endorse the critical argument for the inclusion of probability and statistics in upper secondary school. This finding would be no surprise to many researchers, based on the very nature of official curricula as a genre whose top-down approach comes across as undemocratic (Skovsmose & Valero, 2002). The curriculum is still at technology associated with mathematics education that a critical paradigm attempts to analytically distance itself from (Pais et al., 2012; Valero et al., 2015).

However, that does not stop us from imagining learning environments aiming to engage students in critique (Skovsmose & Borba, 2004). That leads us to **RQ3**:

RQ3 What considerations should there be accounted for to design learning environments compatible with a critical justification for probability and statistics, under high school teaching conditions in Chile?

In Section 4.2 I have suggested three general design principles that guided the construction of the learning environments of this thesis. The critical justification implies the inclusion of examples of uses and applications of probability and statistics in socially relevant issues, and the *exemplary* principle defines the quality of those examples. Based on the educational philosophy of Dewey (1938/2015, 1916/1966) and calls by Skovsmose (e.g. 1994, 2011) to move away from an exercise paradigm to landscapes of investigation, the *inquiry approach* guides the structure of the learning environments. Moreover, since the attempt is for the learning environments to take place under current teaching conditions, the principle of *pragmatism* serves as a ground wire negotiated with the teachers involved.

I proposed these principles with the caveat that some tensions may occur between and within them, which I have addressed in Chapter 9 in light of the classroom experiences. Instead of a strict checklist, design principles should then be viewed as design dimensions to be aware of, and each deserves some reexamination.

Exemplarity can be mistaken as the mere attribute of using selected cases as a teaching strategy to deal with an overwhelming curriculum. However, I have interpreted it as the quality of such cases, focusing efforts onto selecting them properly. Moreover, it must be noted that examples do not speak by themselves, and exemplarity should also be an attribute of the teaching-learning process. For instance, concerning the critical perspective on exemplarity (Negt, 1975), the crowd size estimation task in Chapter 8 was picked to explore how crowd numbers can play a role in mainstreaming an otherwise fringe cause. Upon the lack of reflections in that direction, I asked about it directly, which did not work either. How much can we push it to reflect upon issues we find relevant? Carla describes her position as a teacher:

CARLA: That is the point, like, we do not want people to believe that we are indoctrinating them and saying "this is the best position, I support – or not – the students' march". Instead, we want them to have the tools to say "... at least I know how it works. And I will not be fooled".

In projects such as "Terrible small numbers" (Alrø & Skovsmose, 2002; Alrø et al., 2006), authors accept that the students did not experience the intended exemplarity. Teacher and researchers wanted to make their inquiry exemplary for the issue of reliability and responsibility. However, the scope of reflections, on behalf of the students, does not go beyond the classroom activity. The possible solution thy propose is to offer them different perspectives. After all, part of doing critique is embracing uncertainty by rejecting any dogmatic position, even towards critique itself (Skovsmose, 2008).

An *instrumental* perspective on examplarity presumes that disciplines have elementary problems and questions (Wagenschein, 1956) and thus the quality of an example is its ability to illustrate them. In statistics education, scholars have proposed to focus on big ideas rather than specific methods and procedures, such as data, variation and inference (P. Cobb & McClain, 2004; Ben-Zvi, Gravemeijer, & Ainley, 2018). Concerning the latter, the research programme of informal inference (e.g. Zieffler et al., 2008; Makar & Rubin, 2009) provides a significant opportunity to delineate and examine the fundamental ideas behind statistical inference, beyond its formal procedures, thus alleviating the tension between exemplarity and the curricular coverage.

In Chapter 5, making connections between the juice experiment and the case of Malcolm Collins may be considered to be forced. In the former, we spent a full 90-minute session with Alejandra and her students to explore, define and exercise binomial probabilities to compute the P-value. In the latter, the binomial model was not needed. This discontinuity is not a problem if we consider the big idea to be inference. Hypothesis testing becomes a type of inferential problem in the form of comparison between two models, the comparison between observed data and expected outcomes (Zieffler et al., 2008; Lee et al., 2010), or, in its original meaning, the exercise of distinguishing art from chance (Aburthnot, 1710). In that sense, the connection between both cases is not forced. Still, as with the critical perspective, exemplarity needed to be experienced by a further reflection and institutionalisation of inference in the plenum.

Similarly, in Chapter 8, the big statistical idea at hand is not the construction of confidence intervals, but inference as a type of reasoning that involves the use of data, assumptions of generalisation to a population and an explicit account for the uncertainty involved (Makar & Rubin, 2009). Carla succeeded in introducing the particular case of normal confidence intervals as a procedure, which students followed to make their estimates. Had we made the case exemplary in an explicit way to Makar's framework, students could have made use of additional tools to scrutinise the non-transparent approach of the authorities to estimate the size of the crowd. Students built 95% confidence intervals, how do authorities account for uncertainties when the estimate is given as one number alone?

As for probability, there exist many frameworks attempting to encompass its big ideas in a comprehensive way (Mooney et al., 2014). From a practitioner's perspective, Jones et al. (1997) sum up probabilistic thinking into five constructs: sample space, probability of an event, probability comparisons, conditional probability and independence. As part of his probability literacy framework, Gal (2005) defines five big ideas within the knowledge base dimension: variation, randomness, independence and predictability/uncertainty. The learning environments adhering to the probability curriculum in this thesis are dedicated to risk and decision making, namely in the context of conviction (Chapter 5) and in environmental issues (Chapter 7). Thus the framework proposed by Borovcnik (2017) seems more appropriate, comprising the theoretical character of probability and independence, conditional probability, and concepts building on probabilistic evidence.

Adopting an *inquiry approach* is a broad stance since inquiry-based pedagogy comprises several frameworks for action in mathematics and science education (Artigue & Blomhøj, 2013). I will focus n three main findings after the classroom experiences: its loose definition, authenticity and the guiding role of the teacher.

If a loose definition of inquiry-based teaching is a strategy whereby students work emulating what professionals do, then a fundamental question is: how do statisticians and (mathematical) probabilists work? The inquiry cycle of statisticians has been investigated in the seminal work of Wild and Pfannkuch (1999), roughly summarised as PPDAC (Problem, Plan, Data, Analysis, Conclusion). A more recently proposed cycle, from a broader perspective on data analytics, adds two further steps, namely making decisions and evaluating courses of action (Kazak, Fujita, & Turmo, 2019). I am not aware of any parallel depiction for probabilistic work, other than integrating probabilistic concepts and thinking in statistical inquiries. There are, of course, experimental (e.g. Nilsson, 2014) and problem-led (e.g. Sharma, 2016) approaches to build on probabilistic notions, but further research is needed to define what a probabilistic inquiry cycle may be if it can be conceived at all.

In my opinion, the compass of probabilistic-based inquiry points towards risk and decision making (Borovcnik, 2015, 2017). There are many relevant frameworks enlisting factors of interest in decision making under risk scenarios, including judgements of probability, judgements of impact, coordination of likelihood and impact, the influence of beliefs and values, and the definition of stakeholders (Pratt et al., 2011; Levinson et al., 2012; Radakovic, 2015b). However, a remaining question is whether these elements can define a sequential model or cycle. What is clear is the type of guiding inquiry, namely a scenario where a choice is to be made. In

Bernardo's opinion, this inquiry approach represents a new opportunity for teaching probability:

> BERNARDO: Do I like more statistics or probability? I tell you, easily, I have gone more crazy about statistics than probability. But this vision, this type of exercise, opened my eyes to a sense of probability that goes beyond what the curriculum indicated. And it can be explored differently, as statistics is analysed as well. Because statistics is what normally we apply to this type of problems, this type of discussions …

The type of problems Bernardo refers to are those of social and political relevance. He further exemplifies with the use of statistical information on income distribution and poverty rates in public debates and policymaking. It makes sense, then, that if the critical justification for teaching probability in school is brought up, a path to follow is to include matters of societal risks (Beck, 1992). Still, the extent to which probability calculations inform actual decision making in society remains a pending issue. It is part of the question of authenticity.

The *authenticity* of a task is not a binary feature, nor a one-dimensional spectrum. I have drawn on Palm's (2009) framework developed for word problems to address the authenticity of the inquiries. All four learning environments relate to authentic events and: the case of Malcolm Collins, publications about PSU scores in the media take place every year, the thermoelectric power plant project is already approved, and the students' took part of the march. They vary mostly in the dimensions of *data* and *solution strategies and requirements*. The experiences suggest that tasks with authentic available data can be engaging, and can make smoother connections to real-life use of disciplinary ideas.

However, the experience in Chapter 7 shows that even if a learning environment is openly designed in such a way that data and procedures are fictitious, students can engage in rich reflections anchored in a context that is close to them. Bernardo reflects on the dialectical tension of making the task more faithful by letting students search for relevant data on the power plant project. The complexity of scenarios and the decision-making process would be out of the scope of students reach. Furthermore, we do not know what the true power of mathematics in such process is, that involves a network of power dynamics that goes beyond calculations of likelihood, cost and benefit. In that regard, I find relevant the criticism Valero (1999, p. 22) makes about the notion of the formatting power of mathematics:

> The critical ideology overemphasises the role of mathematics in society. In Latin America, the power structure has lead to a clientelist political system where decisions are made based on personal loyalty of clients to patrons, political convenience, power of conviction through the use of language or violent and physical imposition. In this "rationality", mathematics does not necessarily constitute a formatting power that greatly influences decision making.

The case of "Should we install a thermoelectric power plant?" takes part of this criticism. In order to "didactify" a social problem as input for reflections, the role of mathematics is forced into the learning environment.

The third element of the inquiry approach I would like to revisit is the guiding role of the teacher, specifically regarding mathematical ideas. For example, in Chapter 6, students benefited from the second session, where we exercised the computation of PSU scores, based on the table of conversion, for three different scenarios. Understanding the sorting character of the PSU scale was a necessary input for Miguel to provide alternative interpretations to the graphical interpretation. I have pointed out in Chapter 9 at inquiry-based learning needs not to be a fully unguided learning strategy. Moreover, the experiences show that supporting students and crystallising mathematical ideas can help both to advance the inquiry and to nurture students' reflections.

Pragmatism was meant to be a more or less strict set of restrictions that define a polyhedron wherein we move for the design and development of the learning environments. My aim was not to disturb the teachers' plans and work. However, these conditions can still be flexible in practice. More importantly, it is good to remember that "practical organisation presupposes negotiation" (Skovsmose & Borba, 2004, p. 218).

Time availability can vary. In Chapter 5, I mentioned that we could have used more time to discuss the case of Malcolm Collins. However, a few hours after the "Guilty or innocent?" experience had finished, I was asked to stay with the 11th-grade class. A teacher had called in sick. I decided not to follow on the inquiry without Alejandra being present, and the collection of video and audio would have "violated" the signed consents. We spent that time chatting and solving some mathematics puzzles students offered. Had I been more open, those casual 45 minutes could have been an interesting follow up to the inquiry.

What is considered to be *previous knowledge* is undetermined as well. In the process of designing the "Should we install a thermoelectric power plant?", one of the main inputs was Bernardo's teaching plan for the year, according to which conditional probabilities would have been covered by the time of the intervention. This was not the case. Moreover, conditional probabilities were the upper bound for that group of students that year:

BERNARDO: Then I know they reached conditionality. I do not remember well if we saw the binomial [distribution] (...), let alone random variables. For the simple reason that they were delayed in curricular coverage.

Bernardo realised this just before the implementation of our task, at the moment he made the diagnose test he does at the beginning of each unit (e.g. probability). Skovsmose (2011) argues – conjectures really – that there may be a bias in mathematics education, in that research literature tends to visualise prototypical conditions for teaching and learning, with "ideal classrooms, ideal teachers and ideal students" (p. 19). The notion of previous knowledge in the design should be a consideration for adaptation, not a firmly held assumption.

The *curricular framework* is a guideline, but teachers are the actual "owners" of their class. We held a meeting with Carla right after the first session of "How many people attended the students' march?". I was worried about the prospect of not being able to cover the mathematical notions we were required by the curriculum, especially considering that her students were less than two months away from taking their PSU tests. I have a field note about her reaction: "PSU aside – so we relax". She, more than I, recognised that the possibilities unfolded by letting students come up with ideas that were worth the risk of delaying or deviating from the core curriculum (Penteado, 2001; Skovsmose, 2011). In particular, the possibility to enable students reflection is what I address in the following research question.

RQ4 Can students reflect upon the formatting power of mathematics in learning environments designed after the considerations addressed in **RQ3**?

RQ4 calls for empirical evidence on the potential of the learning environments designed and implemented to engage students in critical reflections about the formatting power of mathematics. In Section 9.2, I have addressed this issue by revisiting each of the classroom experiences in light of their respective frameworks for analysis, which I wrap up now. The formulation of **RQ4** is intended to be addressed both as proof of existence – what *did* happen – and as imagination of potential – what *could* happen (Skovsmose & Borba, 2004).

Did students critically reflect? The answer is a shy "yes".

In "Guilty or innocent?" and "Should we install a thermoelectric power plant?", reflections rest mostly over the mathematical and technological orientation. In both cases, it can be explained from the organisation of the learning environment, in particular, about the time allocated to the joint reflection and the ownership students have in steering the inquiry. That is not to simply say that there was not enough time to discuss, but the time allocated was guided in a way that does not promote critique. In "Guilty or innocent?", the discussion is focused in the three issues pinpointed by (Rosenthal, 2015), which, instead of addressing how probabilistic calculations and interpretations shape our views of evidence and justice, they address whether calculations and interpretations are right or wrong. Similarly, in "Should we install a thermoelectric power plant?" students are asked to present their results and decisions. Having to make a decision, however, provoked critical reflections such as the awareness that small probabilities are not equivalent to full security, or that the worst-case scenario of an environmental disaster is not the one reason to oppose the construction of a power plant.

The statistical tasks provide more evidence for critical reflections taking place. In "How are PSU scores computed?", students explicitly reflect upon how graphical representations shape their and the general public's perceptions of a phenomenon. They are capable of providing an alternative, and opposing interpretation of the political message conveyed in Figure 6.3. In "How many people attended the students' march?", traces of critical reflections are more evident. Students realise that assumptions for expanding their estimations to the whole extent of the march must be negotiated and that the size of their estimates affects the perception of an event where they participated. By casting shadows on the sources of data, a discussion takes place about the divide between statistics-or-lies and constructions-and-calculations (Best, 2005).

Could students critically reflect? Examining the potential of the learning environments in light of the classroom experiences leads to a more daring yes. In Chapter 5, the design of the juice experiment implies several considerations that could trigger a reflection on the fact that numbers do not speak for themselves and what counts as data involves numbers put into context (G. W. Cobb & Moore, 1997). These considerations include the use of a blind experiment to convey validity, and repetition and independence to convey reliability. The main potential of the task lies in the use of P-values as probabilistic evidence. Possible reflections that can take place concern the way a P-value based on a probabilistic model (e.g. binomial) with a frequentist interpretation can shape the notion of evidence about the epistemic probability of guilt of a suspect, i.e. the logical interpretation (Batanero et al., 2016). Echoing Hacking's (Hacking, 1975) philosophical address, I find it essential to distinguish between a probability that represents law-like regularities in repeated chance processes and the epistemic degree of belief with no statistical underpinnings.

In Chapter 6, we revised published statistical information on results of PSU scores as an example of the relation between educational assessments and its associations to categories that define kinds of people, mediating double gestures of social (in)exclusions (Andrade-Molina, 2017; Popkewitz & Lindblad, 2018). Aggregated and categorised statistical information could trigger reflections about bridging objective socio-economic backgrounds and subjective educational foregrounds (Skovsmose, 2005a), as a manifestation of the dialectic between structure and individual (Wedege, 2016). Since PSU scores are given in a normalised relative scale, there are opportunities to reflect upon how the normal distribution not only describes a part of the world but prescribes it as a fabricated normality. Addressing the construction of the PSU scores conversion formula (Equation 6.1) could lead to realising that 30% students left out of the higher education applications process is not scandalous, but tautological news. Graphical representations of statistical information can illustrate the social and political tensions in the construction of statistical messages, and the role of ambiguity and political framing in the collection, treatment and reporting of data, hidden under the veil of objectivity that numbers convey (De Maio, 2007; Greer, 2009).

In Chapter 7, exploring the notion of environmental risk allows reflecting on the possible effects of assuming that risks can be grasped operating probability calculations, often based on a frequentist interpretation for a one-off decision (Borovcnik, 2015). The elusive concept of risk "characterises a peculiar, intermediate state between security and destruction, where the *perception* of threatening risks determines thought and action"(Beck, 2000, p. 212, emphasis in original). Students, from a perspective of local stakeholders of a risk decision, are invited to contribute diverse aspects that shape the complexity of risk. The prime potential reflection in the "Should we install a thermoelectric power plant?" is the very fact that there are strong assumptions in modelling the disaster-event as the intersection of arbitrarily many events. The more events and assumptions of independence between them, the resulting small probability would lead to an arbitrary perception of security.

In Chapter 8, the inferential character of the estimation task can be informative of reflections in all three aspects of inference (Makar & Rubin, 2009). First, the explicit use of data can encourage students to realise that data are constructed and take part in human-made jewels, not solid rocks (Best, 2005). Second, the assumptions of generalisation to a population are (politically) negotiated when situated in context. Third, a formal account of the uncertainty involved in the inference can drive students to ask whether authorities do as well. Above all, students can reflect upon the uses of numbers to shape the perception of adherence to a political cause. Contingent events in Chile make this opportunity more vivid, as the crowd size estimates have become part of the public debate. The possibility to critically reflect about the uses of politically framed statistics is, thus, historically situated, as Carla recognises:

> CARLA: Now, nowadays, I mean, with all that has happened, I believe that now I could feel that reflection. "Yes, it is really important to know a bit of mathematics to be able (...) not to believe so much in what the press tells me". But I do not know if that reflection could have been reached. (...) Perhaps I am wrong.

One research question remains to be addressed. It regards the domain-specificity of reflections that *did* and *could* take place in these four learning environments:

RQ5 What role do probabilistic and statistical notions play in students' critical reflections unveiled in **RQ4**?

In order to approach **RQ5**, I have made use of different domain-specific frameworks for analysis to identify probabilistic and statistical notions (e.g. Rosenthal, 2015; Borovcnik, 2017; Weiland, 2017), more or less entangled with notions of critique in general. I do not intend to display a comprehensive list. Instead, I mention a few aspects that stand out during the classroom experiences as relevant for reflecting on their formatting power, which echo delineations of probabilistic (e.g. Jones et al., 1997) and statistical (Garfield & Ben-Zvi, 2008) big ideas.

The very *meaning of a probability* is called for as a fundamental aspect of both probability thinking and literacy (Jones et al., 1997; Gal, 2005). Probability is a mathematical metaphor, and essentially non-empirical (Spiegelhalter, 2014), and so different meanings of probability arise depending on what the metaphor is used for (Batanero & Díaz, 2007a; Batanero & Borovcnik, 2016; Borovcnik, 2017). Borovcnik and Kapadia (2014) delineate three elemental meanings of probability, namely *a priori* (APT), frequentist (FQT) and subjective (SJT), while Batanero and Borovcnik (2016) consider the logical, propensity and axiomatic meanings as well. In "Guilty or innocent?", collected data in the juice experiment are assessed with a binomial model (APT) in the form of a P-value, whose interpretation relies on an FQT perspective. Moreover, P-values are interpreted in the logical sense, i.e. to assess an "objective" degree of belief on a person's tasting ability or a suspect's innocence. In "Should we install a thermoelectric power plant?", basic computations of probabilities rely on FQT interpretations of historical data to figure out probabilities (Gal, 2005). At the same time, the decision at stake is one-off (Borovcnik, 2015). The ambiguous character of the probability of using poor-quality coal ($\mathbb{P}(C) = 30\%$) illustrates the necessity of digging into the interpretation of probabilities associated with a chance experiment. In "How many people attended the students' march?", in order to account for the uncertainty of the estimation, confidence intervals are constructed based on an APT normal distribution. In contrast, their interpretation relies on repeated sampling experiments (FQT) that are hard to grasp even amongst professionals (Gigerenzer, 2003; Erickson, 2006; White & Gorard, 2017). *Probabilities are critical, in the sense that they merge a variety of non-interchangeable meanings into one measure, shaping our conception of more or less ethereal phenomena into one that can be computable.*

Independence has been pointed out by many scholars as an essential construct of probabilistic thinking (Jones et al., 1997; Mooney et al., 2014; Borovcnik & Kapadia, 2011; Borovcnik, 2017). Two events A and B are independent if and only if the property $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ holds. This mathematical definition cannot be tested with the tools available in the curriculum (e.g. χ^2-test). Nevertheless, its assumption can be discussed to be reasonable and, often, its violation can only be regarded in hindsight. In "Guilty or innocent?", it is a necessary condition to compute the P-value both for the juice experiment and for the case of Malcolm Collins. In the latter, independence between features of the suspect reduces the P-value and thus the conception about his innocence. In "Should we install a thermoelectric power plant?", independence between events that jointly lead to an environmental disaster is necessary for computing its probability. Still, it can be corrected via conditional probabilities if the information is available (see next point about *conditional probabilities*). Moreover, expanding the probability of a disaster in the long term would require the assumption of independence between days, which students intuitively question. Consequently, independence plays a role of mathematising the judgement of the likelihood of an unwanted event, and thus the conception of risk. In "How many people attended the students' march?", independence is a necessary condition for the

Central Limit Theorem, underlying the construction of normal confidence intervals, the one inferential tool available for students within the curricular framework (MINEDUC, 2015a). It becomes a formal requirement to address the uncertainty in their estimations (see the paragraph about *inference*). *The assumption of independence is critical, in the sense of being a requirement for several probabilistic calculations and models that shape parts of the world.*

Conditional probabilities may arise as a solution to a doubtful assumption of independence, given that, if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ does not hold, then the general case $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A)$ is true. Conditional probabilities are asymmetric, i.e. $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$. Moreover, the only way to deduce one from another is to apply Bayes' formula, requiring unconditional probabilities $\mathbb{P}(A)$ or $\mathbb{P}(B)$ based on often unavailable information (Gigerenzer, 2002; White & Gorard, 2017). The conditioned (i.e. non-independent) relation between events can be understood in many ways, namely causal, temporal, or a broader sense of association (Borovcnik, 2012). In "Guilty or innocent?", conditional probabilities and their asymmetry give meaning to the discussion of P-values as conveyors of probabilistic evidence. If H_0 is a null hypothesis and D is a statistic associated to collected data, $P - \text{value} = \mathbb{P}(D|H_0)$, but often mistaken as $\mathbb{P}(H_0|D)$, which would be a more useful input for decision making. Analysing the case of Malcolm Collins, an attachment to the causal interpretation of conditionality led to regard non-causal – but arguably associated – events, namely having a moustache and having a beard, as being independent. Unwanted scenarios of wrongful exoneration (type-I error) and wrongful conviction (type-II error) are asymmetric conditioned events as well. Depending on the context and personal views, these can be interpreted as value-laden risks. In Chapter 5, students argue which risk is worth taking; letting a guilty person free or convicting an innocent person, illustrating the dual nature of risks as both factual and value (Beck, 2000). In "Should we install a thermoelectric power plant?", conditional probabilities may correct, to some extent, a questionable assumption of independence for computing $\mathbb{P}(A \cap B)$ (see Chapter 7). However, as Bernardo points out, the assumption of independence between air quality and the use of coal is highly questionable, but no information is available to address it. *Conditional probabilities are critical, in the sense of mathematising a problematic notion of evidence and clustering different incompatible interpretations of association into one mathematical construct.*

The notion and use of *distributions* are among of the big statistical ideas depicted by Garfield and Ben-Zvi (2008) since the omnipresence of variability is at the core of statistics (G. W. Cobb & Moore, 1997) and "statisticians look at variation through a lens which is distribution" (Wild, 2006, p. 11). Garfield and Ben-Zvi (2008) distinguish between two main types of distribution; empirical (from collected data) and theoretical (probabilistic models). Moreover, these can be bridged by fitting or contrasting empirical data with probabilistic models. In "Guilty or innocent?", the theoretical binomial distribution plays the exemplary role of interpreting collected data as probabilistic evidence. On behalf of the teacher and researcher, it guides the process of constructing data as well. In "How are PSU scores computed?", empirical distributions play the role of sorting and classifying people, by income quintiles, school type and educational performance (PSU test percentiles). The theoretical normal distribution enacts the prescriptive function of mathematical models (Davis & Hersh, 1986), by which the results of a test are transformed to behave in a particular standard way. Overlooking the ordering character of empirical distributions can mislead the interpretation of PSU score trends. In "How many people attended the students' march?", the theoretical normal distribution is somewhat hidden or assumed to hold for the construction of confidence intervals. In this case, the third type of distribution suggested by Garfield and Ben-Zvi (2008) takes place, namely that of sampling distributions that account for the precision and variation of estimations by sampling methods. *Distributions are critical, in the sense of shaping (almost literally) our conceptions about order, variation and uncertainty.*

Inference can be portrayed as a kind of inductive reasoning that comprises three main features: the explicit use of data, claims of the extrapolation to the aggregate going beyond data, and some account for the uncertainty involved in the process (Makar & Rubin, 2009). Although more generic approaches to estimation deal with limited data an extrapolations, dealing with uncertainty in a formal manner is the essence of statistical inference. In "Guilty or innocent?", inference takes the form of hypothesis testing, i.e. the comparison between two models by contrasting collected data against chance models for the null hypotheses (Zieffler et al., 2008; Lee et al., 2010), in order to distinguish art versus chance (Aburthnot, 1710; Hacking, 1975). In "How many people attended the students' march?", the uncertainty involved in inference is dealt using 95% confidence intervals, whose construction relies on APT normal distributions and whose interpretation is based on an FQT perspective. However, statistical inference serves as a counterpoint to the estimates given by authorities as a single approximate number. *Statistical inference is critical, in the sense of providing formal tools to deal with the uncertainty involved in making generalisations about a population, based on an assumed-to-be representative sample.*

Probability can oversimplify the perception of a variety of conflicting meanings into one single metaphorical construct. The puzzling theoretical character of independence is a necessary assumption to make the uncertainty of complex systems mathematically graspable. Conditional probabilities cluster different interpretations of non-independence into one asymmetric construct, seldom possible to reverse, and conveys statements of probabilistic evidence and association. Distributions give sense to phenomena by providing shape to order, variation and uncertainty, using theoretical probabilistic models or arrangements of empirical data. Statistical inference provides formal procedures to grasp the uncertainty involved in the inductive process of using sample data to make generalisations about a population.

When used in the situations referred to in the learning environments of this thesis, these elements shape epistemological claims based on circumstantial evidence, create kinds of people and drive public political opinion, steer judgements of risks that inform complex decision making, and give a sense of objective adherence to a popular cause. These elements become evident input for addressing the formatting power of probability and statistics in the Chilean upper secondary school classrooms, as they *did* or *could* take part in critical reflections during the four classroom experiences of this thesis.

10.2 A return trip along the *problématique*

(How) can we build a coherent path along with a critical perspective of teaching and learning probability and statistics in the Chilean upper secondary school context? Throughout my research project, addressing this overarching questioning has both guided and puzzled me. I now conclude this document by navigating back the three overarching problems fields in the general *problématique* of the critical perspective on probability and statistics education, namely the issues of implementation, possibilities and justification (Niss, 1993), tackling the *how*, the *what* and the *why* in mathematics education (Wedege, 2006).

10.2.1 Implementation, the how

This thesis is illustrative of some tensions that occur when trying to implement a teaching practice aiming to critique in the form of addressing the formatting power of probability and statistics. I proposed three main principles – *exemplarity*, *inquiry approach* and *pragmatism* – that guided the design of four learning environments attempting to provoke critique on behalf of 11th and 12th-grade students in different Chilean classrooms, in close collaboration with their mathematics teachers. By addressing **RQ3**, many tensions arose within and between these principles during the implementation of the tasks. They are redefined as dimensions.

Exemplarity stands for more than the use of examples and their quality; it ought to be experienced. A case can be instrumentally exemplary for probabilistic and statistical big ideas pinpointed in **RQ5**. Critical exemplarity was seldom experienced as discussed regarding **RQ4**. The subjective perspective on exemplarity means to make the example relatable to personal experiences, either by experimenting as in Chapter 5 or by contextualising activities to events, students have participated in, as in Chapter 8. Exemplarity is, therefore, a quality of the teaching practice and future research can go in the direction of the role of teachers in making connections to the bigger picture.

An inquiry approach implies guiding learning with an up-front investigation that students ought to explore and make their own. Decision making under relatable and contingent risk situations (Borovcnik, 2015; Beck, 2000) can be approached through probabilistic ideas. The reconstruction of statistical messages that appeal to students' world can engage them in writing their world (Gutstein, 2006; Weiland, 2017) through statistical inquiry (Wild & Pfannkuch, 1999). A dilemma arises concerning the authenticity of inquiries (Palm, 2009). Authentic approaches for decision making and statistical investigations can overwhelm the school curriculum or eventually not reveal a strong influence of mathematics (Valero, 1999). Adapting the tasks may be approachable for students and effectively allow for critique to take place. Nevertheless, it can miss the point of confronting real-life uses of mathematics in society.

The implementation of school practices that point to the preoccupations of CME can be involved in undesired tensions. I highlight what could be an "opportunity" to investigate students' foregrounds (Skovsmose, 2005a; Wedege, 2016) in Chapter 6 elicited by discourses embedded in statistical information about education assessment. One can fall into the same practices that a critical researcher attempts to be critical of, namely, the double gestures that take place in processes of (in)exclusions (Popkewitz & Lindblad, 2018). That is, by spotting structural aspects of students background that influence foregrounds, the effect could become sharper by perpetuating discourses of fatalism associated with particular kinds of people. This dilemma is much in line with the risk that Valero et al. (2015) identify in some attempts of making different curricula for disenfranchised groups.

Pragmatism does not imply a strict attachment to time and curricular boundaries, which can and should be continuously negotiated with the teacher through *practical organisation*

(Skovsmose & Borba, 2004). Beyond the arranged situation that took place, collaboration with teachers as research *subjects* instead of *objects* can make the investigation go on to a new process of pedagogical imagination.

For example, during the interview, Bernardo told me that he has been implementing the ideas if inquiry in probability in lower secondary school and focusing on students' reflections. He tried out a game of dice, and the student who excelled the most was uncomfortable with her success:

BERNARDO: (...) And this girl said: "I was the only one that won". I told her "Yes. Because you are lucky or skilled". She then said: "But I was the only one that won, and if I go to the casino, I may win too, but everyone else in the casino loses". I told her: "Yes, that is what the business is about; the house always wins". She told me: "Yes, but everyone is happy for me. So they are enthusiastic about me winning and believe they can win, but in reality, they are all losing".

It is not surprising to use games of chance to foster probabilistic ideas in school. However, Bernardo is adjusting his lens towards a student's reflections that go beyond the activity. She is close to understanding that some systems are designed to have a few winners and many losers. This transfer echoes the work of Nobre (1989), whereby students not only make sense of a game of lottery but reflect upon profit as the driving force behind the owners' push for people to play.

The discussion between Bernardo and his student gives an interesting sweet spot within the design dimensions. He is developing probabilistic ideas within the restrains of classroom activity. He is doing so through a personally engaging game, and the dialogue with his student helps to make the case exemplary for broader systemic practices. Moreover, the activity is guided by an up-front inquiry of strategy and exploration.

10.2.2 Possibility, the what

By addressing **RQ4** and **RQ5**, I pinpointed probabilistic and statistical big ideas that play a role in formatting aspects of the world. These are the different meanings of probability, the assumption of independence, conditional probabilities, distributions and inference. I have addressed **RQ4** using Ole Skovsmose's notion of critical epistemology, anchored in his *Towards a Philosophy of Critical Mathematics Education*. For Skovsmose, critique in the classroom takes the form of reflective knowing, which "must address problems and uncertainties connected with transitions between the different types of language game involved in the mathematical modelling process (...) not to solve that problem – this cannot be solved – but in order to create an awareness of the nature of the transitions" (Skovsmose, 1994, p. 111). The point is not to move from one doctrine to another, as critique is free of dogmas and embraces indeterminacy (Skovsmose, 2008, 2014b; Ernest et al., 2016). That is, for the sake of my investigation, it suffices to interpret students' reflections as questioning the assumptions behind probabilistic and statistical procedures and language, and the role of inquiry is to turn those realisations and processes into tangible experiences (Dewey, 1938/2015).

Other scholars point to the more emancipatory and normative character of the critical paradigm, often associated with preoccupations of social justice. Gutstein (2006) calls for writing the world with mathematics, and so does Weiland (2017) with statistics. Inquiries, therefore, are set to make changes, to solve problems of injustice. However, for researchers like (Pais et al., 2012), the promotion of critical mathematics education within the confinements of the school system is in itself contradictory with the foundations of *critique* as a conceived within Critical Theory. To a large extent, this is due to what they call a “domestication” of critique to make it fit the school curriculum. My project is close to what they criticise.

The same possibility problem can be found in the notions of probabilistic and statistical literacy. All of these frameworks have some mention to “the critical”. For Gal (2002), critical questions pertain to the knowledge base, and a critical stance is an attitudinal aspect. Others ground the critical levels of hierarchical models in the interaction between mathematical elements and context (Watson & Callingham, 2003; Sharma, 2016). Kuntze et al. (2017) focuses on common elements of statistical and critical thinking, whereas Weiland (2017) intersects statistical and critical thinking. A necessary step would be to review the literature on probabilistic and statistical literacies and make sense of the different meanings given to critique.

Moreover, the remaining question is whether these ambitious goals can be spotted in classroom practice. For Batanero (2002), the development of literacy is a long-term process. Classroom interaction calls for more local sets of lenses.

Overall, the possibility problem in the critical perspective is subjected to our use of the term *critique* and the assumption of critique being evident in classroom interaction.

10.2.3 Justification, the why

Drawing on **RQ1** and **RQ2**, the power of future citizens to make decisions is usually constrained to the personal scope. Thus the curriculum tends not to position future citizens as decision-makers at a societal level. That is, probability and statistics become essential for all to be responsible consumers. Justifying a mathematics education practice in the name of ideals of critical mathematics education implies a problem of compatibility with the curriculum as a top-down technology of governance. S. E. L. Bello and Traversini (2011) position the inclusion of statistics in the Brazilian curriculum as an implement of governance as well. However, within the concluding remarks, the authors comment that “we do not aim for positioning us against the necessity to govern others and ourselves. The intention is to think how we could have another way of being governed” (p. 868, my translation). The curriculum, in its nature as genre or its contents, may not be the most fitting example that activates the critical justification for teaching and learning mathematics in school. However, it is a point of departure to imagine educational practices that do.

Addressing **RQ1** and **RQ2**, I have pointed that curricula tend to accept the premise of the critical justification, namely that we live in a world surrounded by uncertainties and statistical information. Nonetheless, a fundamental tension can be found within the premise of the critical justification. To what extent are probabilistic and statistical ideas formatting society? Valero (1999) points out that, from a Latin American perspective, structures of power may not be as technocratic to put forward mathematics in action as a concern. This apprehension could be upgraded in the context of globalisation.

Nonetheless, addressing **RQ3** raises that problem. Do courts allow circumstantial probabilistic evidence as in Chapter 5, let alone in Chile, where the jury system is not applied at all? As technical as an issue that is the evaluation of a power plant, are probability calculations playing a significant role as in Chapter 7? On the other hand, Skovsmose (1994) makes clear that by no means the formatting power of mathematics is exclusive. Rather, it means that mathematical rationalities take part in a permanent dialectic with other political, economic and cultural motives. If so, by putting the formatting power of (certain notions of) mathematics forward as part of the justification for its inclusion in the school curriculum, one would be overestimating the relevance of mathematics, another preoccupation of critical mathematics education.

Nonetheless, if the premise is accepted, **RQ4** and **RQ5** show that, to some extent and with lax definitions, it is possible for students to critique probability and statistics in an upper secondary school context. Addressing **RQ3** suggests that it is possible to navigate the tensions of designing and implementing teaching-learning practices that allow it. However, this research project has left some unsolved questions and paradoxes.

10.3 Epilogue: Removing the quotation marks

Throughout the thesis, the titles of the learning environments are given in quotation marks, namely "Guilty or innocent?" (Chapter 5), "How are PSU scores computed?" (Chapter 6), "Should we install a thermoelectric powerplant?" (Chapter 7), and "How many people attended the students' march?" (Chapter 8). The function of the quotation marks is to make clear that the inquiries are taken from real or realistic contexts, but the students are not addressing the questions outside of the classroom. The inquiries belong to the learning environment confined to the school context.

Alejandra's 11th-grade students do not and cannot decide whether a suspect is guilty or innocent; they are underage individuals in a country whose trials are not done by juries, discussing a case from the United States in the 1960s. Alejandra's 12th-grade class had the opportunity to answer the inquiry of how PSU scores are computed; they do evaluate a meaningful politically-framed graph and constructed their own versions. Nonetheless, their versions, interpretations and opinions on the graphical representation do not leave the classroom. Bernardo's students are confronted with a decision about the instalment of a power plant in their province. They calculate, they deliberate, they decide. However, they do not have an actual voice; they did not raise their concerns to the authorities. The powerplant's construction has been given the green light. Carla's students made their own estimates for the attendees in the students' march. They discussed their findings, evaluated sources of data, negotiated assumptions for inference and appreciated different sources of variation. However, their answer to the inquiry stays with them.

According to Pais et al. (2012), the natural(ised) constraints of the classroom as a micro-society and the confinement of the school system to the mechanisms of capitalism make the emancipatory ideals of critical mathematics education impossible to achieve. However, they call for acting after the Pascalian maxim of being pessimistic in theory, but optimistic in practice. Pascal's wager, a foundational problem of decision making under uncertainty, is often misunderstood as believing in God for the sake of salvation. Pascal wagers on acting as if one believed in God and "would add: because, if they do so, they will in due course come to believe" (Hacking, 1975, p. 71). Despite the tensions illustrated in this thesis, I do research hoping to believe that the ideals of critical mathematics education can be lived up in the classroom. Teachers understand that it is part of a slow process:

CARLA: I mean, despite external people coming (to our school) and, maybe, they giving us feedback, one does not always feel that it will lead to change. Because I imagine that investigations always have that, like, one investigates (...) and finds a result. And, well, perhaps the result will not always lead to a change. Even so, one can say "oh, I found this, and this is important". I think I do not feel that while I am teaching.

I hope that elements of this research project are important. I hope future students and citizens will have a chance to remove the quotation marks.

Bibliography

Abujatum, J. (2018, October). *Requisitos de ingreso a las escuelas matrices de oficiales delas Fuerzas Armadas, de Orden y Seguridad* [Requirements for admission to main schools of officers in the Arms, Order and Security Forces]. Valparaíso, Chile: Biblioteca del Congreso Nacional de Chile. Retrieved from https://www.bcn.cl

Aburthnot, J. (1710). An argument for divine providence, taken from the constant regularity observ'd in the births of both sexes. *Philosophical Transactions*, *27*(328), 186–190. doi: 10.1098/rstl.1710.0011

Achenwall, G. (1749). *Abriss der neuesten Staatswissenschaft der vornehmsten Europäischen Reiche und Republicken* [Outline of the latest political science of the most distinguished European empires and republics]. Göttingen, Germany: Schmidt.

Agencia de Calidad de la Educación. (2016, 2017). *Base de Datos de la Agencia de Calidad de la Educación* [Database of the Agency for Quality in Education]. Santiago, Chile. Retrieved from http://informacionestadistica.agenciaeducacion.cl

Agencia de Calidad de la Educación. (2019). *Descripción de grupos socioeconómicos* [Description of socioeconomic groups]. Santiago, Chile. Retrieved from https://www.simce.cl/ficha2019/

Aguilar, M. S., & Zavaleta, J. G. M. (2012). On the links between mathematics education and democracy: A literature review. *Pythagoras*, *33*(2), 1–15. doi: 10.4102/pythagoras.v33i2.164

Ainley, J., & Pratt, D. (2014). Chance re-encounters: 'computers in probability education' revisited. In D. Frischemeier, P. Fischer, R. Hochmuth, & P. Bender (Eds.), *Using tools for learning mathematics and statistics* (pp. 165–177). New York, NY: Springer.

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, *32*(1), 23–38.

Alrø, H., Blomhøj, M., Bødtkjer, H., & Skovsmose, O. (2006). Farlig små tal - almendannelse i et risikosamfund [Terrible small numbers - general education in a risk society]. In O. Skovsmose & M. Blomhøj (Eds.), *Kunne det tænkes? - om matematiklæring* (pp. 24–39). Copenhagen, Denmark: Malling Beck.

Alrø, H., & Skovsmose, O. (2002). *Dialogue and learning in mathematics education: Intention, reflection, critique*. Dordrecht, The Netherlands: Kluwer.

Alrø, H., Skovsmose, O., & Valero, P. (2009). Inter-viewing foregrounds: Students' motives for learning in a multicultural setting. In M. César & K. Kumpulainen (Eds.), *Social interactions in multicultural settings* (pp. 13–37). Rotterdam, the Netherlands: Sense. doi: 10.1163/9789087907174_003

Andersen, A. S., & Kjeldsen, T. H. (2015). A critical review of the key concepts in PPL. In

A. S. Andersen & S. Heilesen (Eds.), *The Roskilde model: Problem-oriented learning and project work* (pp. 17–35). Heidelberg, Germany: Springer.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, *25*(4), 5–11. doi: 10.3102/0013189X025004005

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... Wittrock, M. C. (2001). The cognitive process dimension. In L. W. Anderson et al. (Eds.), *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Abridged ed., pp. 63–92). New York, NY: Longman.

Andrade-Molina, M. (2017). The adventure of the deceitful numbers. *Journal of Pedagogy*, *8*(2), 9–25. doi: 10.1515/jped-2017-0007

Antivilo, A., Contreras, P., & Hernández, J. (2015). *Tratamiento de puntajes* [Treatment of scores]. Santiago, Chile: Unidad de Estudios e Investigación. Departamento de Evaluación, Medición y Registro Educacional (DEMRE). Universidad de Chile. Retrieved from `http://docplayer.es/23983949-Tratamiento-de-puntajes.html`

Araneda, A. M., del Pino, G., Estrella, S., Icaza, G., & San Martín, E. (2011). *Recomendaciones para el currículum escolar del eje datos y probabilidad* [Recommendations for the school curriculum on the axis of data and chance]. Santiago, Chile: Sociedad Chilena de Estadística.

Artigue, M., & Blomhøj, M. (2013). Conceptualizing inquiry-based education in mathematics. *ZDM - International Journal on Mathematics Education*, *45*(6), 797–810. doi: 10.1007/s11858-013-0506-6

Atweh, B. (2004). Understanding for changing and changing for understanding: Praxis between practice and theory through action research in mathematics education. In P. Valero & R. Zevenbergen (Eds.), *Researching the socio-political dimensions of mathematics education: Issues of power in theory and methodology* (pp. 187–206). Dordrecht, The Netherlands: Kluwer.

Bakker, A., & van Eerde, D. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 429–466). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-017-9181-6_16

Barbosa, J. C. (2006). Mathematical modelling in classroom: A socio-critical and discursive perspective. *Zentralblatt für Didaktik der Mathematik*, *38*(3), 293–301. doi: 10.1007/BF02652812

Batanero, C. (2002). Discussion: The role of models in understanding and improving statistical literacy. *International Statistical Review*, *70*, 37–40.

Batanero, C. (2005). Significados de la probabilidad en educación secundaria [Meanings of probability in secondary education]. *Revista Latinoamericana de Investigación en Matemática Educativa*, *8*(3), 247–263.

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam, The Netherlands: Sense.

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., & Sánchez, E. (2016). Research on Teaching and Learning Probability. In *Research on Teaching and Learning Probability. ICME-13 Topical Surveys* (pp. 1–33). Cham, The Netherlands: Springer. doi: 10.1007/978-3-319

-31625-3_1

Batanero, C., & Díaz, C. (2007a). Meaning and understanding of mathematics. the case of probability. In J. P. Van Bendegen & K. François (Eds.), *Philosophical dimensions in mathematics education* (pp. 107–127). New York, NY: Springer.

Batanero, C., & Díaz, C. (2007b). Probabilidad, grado de creencia y proceso de aprendizaje [Probability, degree of belief and learning process]. In *XIII jornadas nacionales de enseñanza y aprendizaje de las matemáticas* (Vol. 13). Granada, Spain: Federación Española de Profesores de Enseñanza de las Matemáticas. Retrieved from `http://www.ugr.es/{~}batanero/pages/ARTICULOS/PonenciaJAEM.pdf`

Batanero, C., Fernandes, J. A., & Contreras, J. M. (2009). Un análisis semiótico del problema de Monty Hall e implicaciones didácticas [A semiotic analysis of the Monty Hall problem and didactical implications]. *Suma: Revista sobre Enseñanza y Aprendizaje de las Matemáticas*, *62*, 11–18.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In *Exploring probability in school* (pp. 15–37). New York, NY: Springer. doi: 10.1007/0-387-24530-8_2

Batanero, C., & Sanchez, E. (2005). What is the nature of high school students' conceptions and misconceptions about probability? In *Exploring probability in school* (pp. 241–266). New York, NY: Springer. doi: 10.1007/0-387-24530-8_11

Beck, U. (1992). *Risk society: Towards a new modernity*. London, UK: Sage.

Beck, U. (2000). Risk society revisited: Theory, politics and research programmes. In B. Adam, U. Beck, & J. Van Loon (Eds.), *Risk society and beyond: Critical issues for social theory* (pp. 211–229). London, UK: Sage.

Beck, U., Giddens, A., & Lash, S. (1994). *Reflexive modernization: Politics, tradition and aesthetics in the modern social order*. Cambridge, UK: Polity.

Bello, A. (1843). Discurso pronunciado por el Sr. Rector de la Universidad, D. Andrés Bello, en la instalación de este cuerpo el día 17 de setiembre de 1843 [Speech pronounced by Mr. Rector of the University, Andrés Bello, on the inauguration of this body the day September 17th of 1843]. *Anales de la Universidad de Chile*, *1*, 139–152.

Bello, S. E. L., & Traversini, C. S. (2011). Saber estatístico e sua curricularização para o governamento de todos e de cada um. *Bolema - Mathematics Education Bulletin*, *24*(40), 855–871.

Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In D. Ben-Zvi, K. Makar, & J. B. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 473–502). Cham, The Netherlands: Springer. doi: 10.1007/978-3-319-66195-7_16

Ben-Zvi, D., Makar, K., & Garfield, J. B. (Eds.). (2018). *International Handbook of Research in Statistics Education*. Cham, The Netherlands: Springer. doi: 10.1007/978-3-319-66195-7

Ben-Zvi, D., & Sharett-Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about distributions: A collection of recent research studies. proceedings of the fourth international research forum for statistical reasoning, thinking and literacy (SRTL-4), University of Auckland (New Zealand).* Brisbane, Australia: University of Queensland.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. Garden City, NY: Doubleday.

Bertsekas, D. P., & Tsitsiklis, J. N. (2000). *Introduction to probability*. Cambridge, MA: Massachusetts Institute of Technology.

Best, J. (2005). Lies, calculations and constructions: Beyond "How to lie with statistics". *Statistical Science*, *20*(3), 210–214. doi: 10.1214/088342305000000232

Bhaskar, R. A. (1986). *Scientific realism and human emancipation*. London, UK: Verso.

Design-Based Research Collective. (2003). Design-based research: An emerging paradgim for educational inquiry. *Educational Researcher*, *32*(1), 5–8.

Revista de Educación. (2015, March 25). Chile, la OCDE y la desigualdad en educación: "El sistema escolar en Chile no funciona bien" [Chile, OECD and inequality in education: "Chile's school system does not work well"]. *Revista de Educación*. Retrieved from http://www.revistadeeducacion.cl/chile-la-ocde-y-la-desigualdad-en-educacion-el-sistema-escolar-en-chile-no-funciona-bien

Bidwell, J. K., & Clason, R. G. (Eds.). (1970). *Readings in the history of mathematics education*. Washington, D.C.: National Council of Teachers of Mathematics.

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, *21*(1), 33–46. doi: 10.1007/s11092-008-9064-9

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York, NY: Academic press.

Binet, A. (1899). La pédagogie scientifique. *L'Enseignement Mathématique*, *1*, 29–38.

Blomhøj, M. (2006). Konstruktion af episoder som forskningsmetode - læringsmuligheder i IT-støttet matematikundervisning [Construction of episodes as a research method - learning possibilities in IT-supported mathematics teaching]. In O. Skovsmose & M. Blomhøj (Eds.), *Kunne det tænkes? - om matematiklæring* (pp. 228–254). Copenhagen, Denmark: Malling Beck.

Blomhøj, M. (2013). Hvad er undersøgende matematikundervisning - og virker den? [What is inquiry-based mathematics teaching - and does it work?] In M. Wahl Andersen & P. Weng (Eds.), *Håndbog om matematik i grundskolen. læring, undervisning og vejledning*. Copenhagen, Denmark: Dansk Psykologisk Forlag.

Blomhøj, M. (2016). *Fagdidaktik i matematik* [Didactics of mathematics]. Frederiksberg, Denmark: Frydelund.

Blomhøj, M., & Kjeldsen, T. H. (2011). Students' Reflections in Mathematical Modelling Projects. In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling* (pp. 385–395). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-007-0910-2_38

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects - State, trends and issues in mathematics instruction. *Educational Studies in Mathematics*, *22*(1), 37–68.

Boaler, J. (1997). *Experiencing school mathematics*. Buckingham, UK: Open University Press.

Borovcnik, M. (2006). Probabilistic and statistical thinking. In M. Bosch (Ed.), *European Research in Mathematics Education IV: Proceedings of the Fourth Congress of the Euro-*

pean Society for Research in Mathematics Education (CERME 4, February 17 - 21, 2005) (pp. 485–506). Sant Feliu de Guíxols, Spain: FUNDEMI IQS – Universitat Ramon Llull and ERME.

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics - Challenges for teaching and teacher education. a joint ICMI/IASE study: The 18th ICMI study* (pp. 71–83). Dordrecht, the Netherlands: Springer. doi: 10.1007/978-94-007-1131-0_11

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didáctica de la Matemática*, *2*, 5–27.

Borovcnik, M. (2015). Risk and decision making: The "logic" of probability. *The Mathematics Enthusiast*, *12*(1,2&3), 113–139.

Borovcnik, M. (2017). Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa*, *18*(3), 1491–1516.

Borovcnik, M., & Kapadia, R. (2009). Research and developments in probability education. *International Electronic Journal of Mathematics Education*, *4*(3), 111–130.

Borovcnik, M., & Kapadia, R. (2011). Modelling in probability and statistics. In J. Maasz & J. O'Donoghue (Eds.), *Real-world problems for secondary school mathematics students* (pp. 1–43). Rotterdam, the Netherlands: Sense. doi: 10.1007/978-94-6091-543-7_1

Borovcnik, M., & Kapadia, R. (2014). A historical and philosophical perspective on probability. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 7–34). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-007-7155-0_2

Bourdieu, P. (1980). *Le sens practique* [The logic of practice]. Paris, France: Les Éditions de Minuit.

Brantlinger, A. (2014). Critical mathematics discourse in a high school classroom: Examining patterns of student engagement and resistance. *Educational Studies in Mathematics*, *85*(2), 201–220. doi: 10.1007/s10649-013-9506-2

Brown, W. (2015). *Undoing the demos: Neoliberalism's stealth revolution*. Cambridge, MA: MIT Press.

Brusseau, G. (2002). *Theory of didactical situations in mathematics* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Eds.). Dordrecht, the Netherlands: Kluwer. doi: 10.1007/0-306-47211-2

Bryman, A. (2015). *Social Research Methods* (5th ed.). Oxford, UK: Oxford University Press.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., ... Landes, N. (2006). *The BCSC 5E instructional model. Origins, effectiveness and applications.* Colorado Springs, CO: Biological Sciences Curriculum Study. Retrieved from `http://go.cengage.com/uploads/f5f2e1f839cc8154549259131f39392f_2_7287.pdf`

Børne- og Undervisningsministeriet. (2017). *Matematik B - HTX* [Mathematics B, HTX]. Copenhagen, Denmark: Styrelsen for Undervisning of Kvalitet, Børne- og Undervisningsministeriet.

Børne- og Undervisningsministeriet. (2018a). *Matematik A/B/C, stx: Vejledning* [Mathematics A/B/C, STX: Guidelines]. Copenhagen, Denmark: Styrelsen for Undervisning of

Kvalitet, Børne- og Undervisningsministeriet.

Børne- og Undervisningsministeriet. (2018b). *Matematik A/B, htx: Vejledning* [Mathematics A/B, HTX: Guidelines]. Copenhagen, Denmark: Styrelsen for Undervisning of Kvalitet, Børne- og Undervisningsministeriet.

Børne- og Undervisningsministeriet. (2018c). *Matematik B/C, hf: Vejledning* [Mathematics B/C, HF: Guidelines]. Copenhagen, Denmark: Styrelsen for Undervisning of Kvalitet, Børne- og Undervisningsministeriet.

Cansado, E. (1955). Enseñanza de la estadística [Teaching of statistics]. *Anales de la Universidad de Chile*, *113*(97-98), 121–125.

Carnap, R. (1950). *Logical foundations of probability*. Chicago, IL: University of Chicago Press.

Castillo, V., & Labrín, S. (2020, March 8). Marcha 8m: experto estima en 800 mil las mujeres asistentes a movilización [8m march: expert estimates at 800 thousand the women attending mobilisation]. *La Tercera*. Retrieved from https://www.latercera.com/nacional/noticia/marcha-8m-experto-estima-en-800-mil-las-mujeres-asistentes-a-movilizacion/KRJAH7ZXWRCCRK6Y65X3ALFZZU/

Cavieres, E. A. (2011). The class and culture-based exclusion of the chilean neoliberal educational reform. *Educational Studies*, *47*(2), 111–132. doi: 10.1080/00131946.2011.554588

Chellougui, F., Thu, H. N. T., & Winsløw, C. (2016). Language diversity in research on language diversity in mathematics education. In R. Barwell (Ed.), *Mathematics education and language diversity. new icmi study series* (pp. 263–277). Cham, the Netherlands: Springer. doi: 10.1007/978-3-319-14511-2_14

Chernoff, E. J., & Sriraman, B. (2014a). Introduction to Probabilistic Thinking: Presenting Plural Perspectives. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. xv–xviii). Dordrecht, the Netherlands: Springer.

Chernoff, E. J., & Sriraman, B. (Eds.). (2014b). *Probabilistic Thinking: Presenting Plural Perspectives*. Dordrecht, the Netherlands: Springer. doi: 10.1007/978-94-007-7155-0

Cirillo, P., & Taleb, N. N. (2016). On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications*, *452*, 29–45.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823. doi: 10.1080/00029890.1997.11990723

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical analysis. *Mathematical Thinking and Learning*, *1*, 5–43.

Cobb, P., & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 375–395). Dordrecht, The Netherlands: Springer.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, *21*(1), 1–78.

Comisión Presidencial Pensiones. (2015, September). *Informe Final de la Comisión Asesora Presidencial sobre el Sistema de Pensiones* [Final Report]. Retrieved from http://www.comision-pensiones.cl/Informe_final_CP_2015.pdf

Cuevas, J. (2012). Panorama actual de los estándares educativos en estocástica. *Revista Digital Matemática, Educación e Internet*, *12*(2), 1–12. doi: 10.18845/rdmei.v12i2.1672

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*(5), 382–393. doi: 10.2307/749086

Dahl, B., & Stedøy, I. M. (2004). A Nordic community: Ideas of education and democracy in mathematics. In I. M. Stedøy (Ed.), *Mathematics education - the Nordic way* (pp. 1–10). Trondheim, Norway: TAPIR.

Daly, M. (2005). Gender mainstreaming in theory and practice. *Social Politics: International Studies in Gender, State & Society*, *12*(3), 433–450. doi: 10.1093/sp/jxi023

D'Ambrosio, U. (1994). Cultural framing of mathematics teaching and learning. In R. Biehler, R. W. Scholz, R. Strässer, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 443–455). Dordrecht, the Netherlands: Springer.

David, F. N. (1962). *Games, gods and gambling*. London, UK: Griffin.

Davis, P. J., & Hersh, R. (1986). *Descartes' dream: The world according to mathematics*. Brighton, UK: Harvester.

De Maio, F. (2007). Estimating mortality in war-time Iraq: A controversial survey with important lessons for students. *Teaching Statistics*, *29*(2), 34–37. doi: 10.1111/j.1467-9639.2007.00268.x

de Abreu, G. (2000). Relationships between macro and micro socio-cultural contexts: Implications for the study of interactions in the mathematics classroom. *Educational Studies in Mathematics*, *41*(1), 1–29. doi: 10.1023/A:1003875728720

De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2002). Improper use of linear reasoning: An in-depth study of the nature and the irresistibility of secondary school students' errors. *Educational Studies in Mathematics*, *50*(3), 311–334.

de Carvalho, J. B. P. (2014). Mathematics Education in Latin America. In A. Karp & G. Schubring (Eds.), *Handbook on the history of mathematics education* (pp. 335–359). New York: Springer. doi: 10.1007/978-1-4614-9155-2

Desrosières, A. (1991). How to make things which hold together: Social science, statistics and the State. In P. Wagner, B. Wittrock, & R. Whitley (Eds.), *Discourses on society* (Vol. 15, pp. 195–218). Dordrecht, the Netherlands: Kluwer.

Dewey, J. (1966). *Democracy and education*. New York, NY: Free Press. (Original work published 1916)

Dewey, J. (2015). *Experience and education* (Kindle ed.). New York, NY: Free Press. (Original work published 1938)

Diaper, G. (1990). The Hawthorne Effect: a fresh examination. *Educational Studies*, *16*(3). doi: 10.1080/0305569900160305

diSessa, A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences*, *13*(1), 77–103.

Donnelly, M. (1998). From political arithmetic to social statistics: How some nineteenth-century roots of the social sciences were implanted. In J. Heilbron, L. Magnusson, & B. Wittrock (Eds.), *The rise of the social sciences and the formation of modernity* (pp. 225–239). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-011-5528-1_9

Eisenhart, M. A. (1991). Conceptual frameworks for research circa 1991: Ideas from a cultural anthropologist; implications for mathematics education researchers. In *Proceedings of* *the 13th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 202–219). Blacksburg, VA.

Ekstrøm, C. T., Ernst, H., & Brockhoff, P. B. (2017). *Statistik i gymnasiet* [Statistics in high school]. Copenhagen, Denmark: Børne- og Undervisningsministeriet.

Elicer, R. (2019). The role of decision-making in the legitimation of probability and statistics in Chilean upper secondary school curriculum. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME 11, February 6 – 10, 2019)* (pp. 988–989). Utrecht, The Netherlands: Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Elicer, R. (2020). Meanings of decision-making in probability and statistics: Comparing Chilean and Danish upper secondary school curricula. *Nordic Studies in Mathematics Education*, *25*(1), 25–44.

Elicer, R., & Carrasco, E. (2016). Probabilidad condicional como herramienta para la toma de decisiones [Conditional probability as a tool for decision making]. In *Acta Latinoamericana de Matemática Educativa 29* (pp. 157–165). México DF, México: Comité Latinoamericano de Matemática Educativa A.C.

Elicer, R., & Carrasco, E. (2017). Conditional probability as a decision-making tool: A didactic sequence. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME 10, February 1 – 5, 2017)* (pp. 748–755). Dublin, Ireland: DCU Institute of Education and ERME.

Ellis, K. M. (1995). *The Monty Hall problem.* Retrieved from `http://www.montyhallproblem.com`

Erickson, T. (2006). Using simulation to learn about inference. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education. proceedings of the seventh international conference on teaching statistics, Salvador, Brazil.* Voorburg, the Netherlands: International Statistical Institute.

Ernest, P. (2010). The scope and limits of critical mathematics education. In H. Alrø, O. Ravn, & P. Valero (Eds.), *Critical mathematics education: Past, present and future* (pp. 65–87). Rotterdam, the Netherlands: Sense.

Ernest, P., Skovsmose, O., van Bendegem, J. P., Bicudo, M., Miarka, R., Kvasz, L., & Moeller, R. (2016). The philosophy of mathematics education. In *The Philosophy of Mathematics Education. ICME-13 Topical Surveys* (pp. 1–26). Cham, The Netherlands: Springer. doi: 10.1007/978-3-319-40569-8_1

Fairclough, N. (1992). Discourse and text: Linguistic and intertextual analysis within discourse analysis. *Discourse & Society*, *3*(2), 193–217.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. London, UK: Routledge.

Fairclough, N. (2010). A dialectical-relational approach to critical discourse analysis in social research. In N. Fairclough (Ed.), *Critical discourse analysis: The critical study of language* (2nd ed., pp. 230–254). Harlow, UK: Routledge.

Fajardo, M. (2018, August 29). Combinación fatal en Quintero: mezcla de contaminantes, ausencia de normativas y lobby empresarial [Fatal combination in Quintero: a mix of polluting agents, absence of norms and corporate lobby]. *El Mostrador*.

Retrieved from https://www.elmostrador.cl/cultura/2018/08/29/combinacion-fatal-mezclas-de-contaminantes-ausencia-de-normativas-y-lobby-empresarial

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York, NY: John Wiley & Sons.

Fenton, N., Neil, M., & Thieme, N. (2018). Lawnmowers versus terrorists. *Significance*, *15*(1), 12–13. doi: 10.1111/j.1740-9713.2018.01104.x

Fischbein, E. (1975). *The intuitive source of probability thinking in children*. Dordrecht, the Netherlands: Reidel.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, *15*(1), 1–24. doi: 10.1007/BF00380436

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105.

Frankenstein, M. (1989). *Relearning mathematics: a different third r – radical maths*. London, UK: Free Association Books.

Frankenstein, M. (1994). Understanding the politics of mathematical knowledge as an integral part of becoming critically numerate. *Radical Statistics*, *56*, 22–40.

Franklin, C., & Garfield, J. B. (2006). The guidelines for assessment and instruction in statistics education (GAISE) project: Developing statistics education guidelines for pre K-12 and college courses. In G. F. Burill (Ed.), *Thinking and reasoning about data and chance: Sixty-eight NCTM yearbook.* Reston, VA: National Council of Teachers of Mathematics.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. Alexandria, VA. doi: 10.3928/01484834-20140325-01

Freire, P. (1993). *Pedagogy of the oppressed*. London, UK: Penguin Books. (Original work published 1970)

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of the Political Economy*, *56*(4), 279–304.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education*, *32*(2), 124. doi: 10.2307/749671

Fábrega, J. (2017, January 4). Resultados en la PSU de liceos "emblemáticos": Ninguna de las anteriores [PSU results of "emblematic" high schools: None of the above]. *Centro de Investigación Periodística*. Retrieved from https://ciperchile.cl/2017/01/04/resultados-en-la-psu-de-liceos-emblematicos-ninguna-de-las-anteriores

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25. doi: 10.1111/j.1751-5823.2002.tb00336.x

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In J. B. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht, the Netherlands: Kluwer.

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–63). New York, NY: Springer. doi: 10.1007/0-387-24530-8_3

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22–38.

Garfield, J. B., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372–396. doi: 10.1111/j.1751-5823.2007.00029.x

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer. doi: 10.1007/978-1-4020-8383-9

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you.* New York, NY: Simon & Schuster.

Gigerenzer, G. (2003). *Reckoning with risk.* London, UK: Penguin.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*(1), 20–29.

Gjørup, H., Jensen, N., Hedemann Jensen, P., Kristensen, L., Nielsen, O., Petersen, E., ... Aarkrog, A. (1982). *Radioactive contamination of Danish territory after core-melt accidents at the Barsebäck power plant* (Tech. Rep. Nos. RISO-R–462). Roskilde, Denmark: Risø National Laboratory.

Goos, M. (2018). A review of research on understanding and promoting students' mathematical thinking published in Educational Studies in Mathematics 2014-2018. In D. M. Gómez (Ed.), *Proceedings of the First PME Regional Conference: South America* (pp. 1–16). Rancagua, Chile: PME.

Gorard, S., & White, P. (2017). Still against inferential statistics: Rejoinder to Nicholson and Ridgway. *Statistics Education Research Journal*, *16*(1), 74–79.

Gøtze, P. (2019). How do students make use of their mathematical knowing when they read the world? In J. Subramanian (Ed.), *Proceedings of the Tenth International Mathematics Education and Society Conference* (pp. 425–434). Hyderabad, India: MES10.

Gould, S. J. (1994, November 28). Curveball. *The New Yorker*.

Greer, B. (2009). Estimating Iraqi deaths: A case study with implications for mathematics education. *ZDM Mathematics Education*, *41*(1-2), 105–116. doi: 10.1007/s11858-008-0147-3

Greer, B. (2014). Commentary on perspective II: Psychology. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 299–309). Dordrecht, the Netherlands: Springer. doi: 10.1007/978-94-007-7155-0_16

Greer, B., & Mukhopadhyay, S. (2005). Teaching and learning the mathematization of uncertainty: Historical, cultural, social and political contexts. In *Exploring probability in school: Challenges for teaching and learning* (pp. 297–324). New York, NY: Springer. doi: 10.1007/0-387-24530-8_13

Groth, R. E. (2015). Research commentary:boundaries of mathematics education and statistics education communities of practice. *Journal for Research in Mathematics Education*, *46*(1), 4-16.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.

Gürbüz, R., Erdem, E., & Fırat, S. (2014). The effect of activity-based teaching on remedying the probability-related misconceptions: A cross-age comparison. *Creative Education*, *5*(1), 18–30. doi: 10.4236/ce.2014.51006

Gutiérrez, R. (2008). A "gap-gazing" fetish in mathematics education? problematizing research on the achievement gap. *Journal for Research in Mathematics Education*, *39*(4), 357–364.

Gutstein, E. (2006). *Reading and writing the world with mathematics*. New York, NY: Routledge.

Gómez, D. M. (Ed.). (2018). *Proceedings of the First PME Regional Conference: South America*. Rancagua, Chile: PME.

Hacking, I. (1975). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge, UK: Cambridge University Press.

Hacking, I. (1990). *The taming of chance*. Cambridge, UK: Cambridge University Press.

Halley, E. (1693a). An estimate of the degrees of mortality of mankind drawn from curious tables of births and funerals at the city of Breslaw. *Philosophical Transactions*, *17*(196). doi: 10.1098/rstl.1693.0007

Halley, E. (1693b). Some further considerations on the Breslaw bills of mortality. *Philosophical Transactions*, *17*(198). doi: 10.1098/rstl.1693.0014

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, *39*(5), 832–845. doi: 10.1037/0022-3514.39.5.832

Hansson, S. O. (2009). Risk and safety in technology. In A. Meijers (Ed.), *Handbook of the philosophy of science* (pp. 1069–1102). Oxford, UK: Elsevier.

Hauge, K. H., Gøtze, P., Hansen, R., & Steffensen, L. (2017). Categories of critical mathematics based reflections on climate change. In A. Chronaki (Ed.), *Proceedings of the Ninth International Mathematics Education and Society Conference* (Vol. 2, pp. 522–532). Volos, Greece: University of Thessaly Press and MES9.

Herschel, J. (1850). Quetelet on probabilities. *Edinburgh Review*, *92*, 1–57.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., ... Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, *25*(4), 12–21. doi: 10.3102/0013189X025004012

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107. doi: 10.1080/00461520701263368

Huff, D. (1954). *How to lie with statistics*. New York, NY: Norton.

Hurlbert, M., J. Krishnaswamy, E. Davin, F.X. Johnson, C.F. Mena, J. Morton, S. Myeong, D. Viner, K. Warner, A. Wreford, S. Zakieldeen, Z. Zommers, 2019: Risk Management and Decision making in Relation to Sustainable Development. In: *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems* [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D.C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J.

Malley, (eds.)]. In press.

Illeris, K. (2002). Experiential learning. In *The three dimensions of learning: Contemporary learning theory in the tension field between the cognitive, the emotional and the social.* Frederiksberg, Denmark: Roskilde University Press.

Jablonka, E. (2014). Critical thinking in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 121–125). Dordrecht, the Netherlands: Springer.

Jara, C. (2020, April 8). Quintero-Puchuncaví alcanza peak de contaminación en plena crisis por coronavirus [Quintero-Puchuncaví reaches pollution peak in the midst of coronavirus crisis]. *El Desconcierto*. Retrieved from `https://www.eldesconcierto.cl/2020/04/08/quintero-puchuncavi-alcanza-peak-de-contaminacion-en-plena-crisis-por-coronavirus`

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young chiedren's thinking in probability. *Educational Studies in Mathematics*, *32*(2), 101–125. doi: 10.1023/A:1002981520728

Jones, G. A., & Thornton, C. (2005). An overview of research into the teaching and learning of probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 65–92). New York, NY: Springer. doi: 10.1007/0-387-24530-8_4

Joram, E., Resnick, L. B., & Gabriele, A. J. (1995). Numeracy as cultural practice: An examination of numbers in magazines for children, teenagers, and adults. *Journal for Research in Mathematics Education*, *26*(4), 346–361.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgement under uncertainty: Heuristics and biases*. New York, USA: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292. doi: 10.2307/1914185

Kaiser, G., Blomhøj, M., & Sriraman, B. (2006). Towards a didactical theory for mathematical modelling. *ZDM*, *38*(2), 82-85. doi: 10.1007/BF02655882

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954–2965. doi: 10.1111/jan.13031

Kazak, S., Fujita, T., & Turmo, M. (2019). Building up students' data analytics skills to solve real world problems. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME 11, February 6 – 10, 2019)* (pp. 930–937). Utrecht, The Netherlands: Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Keynes, J. M. (1973). The general theory of employment. In J. M. Keynes (Ed.), *The general theory and after: Part ii defence and development: Vol. xiv. the collected writings of j.m. keynes* (pp. 190–123). London,UK: Macmillan for the Royal Economic Society.

Kilpatrick, J. (1992). A history of research in mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 3–38). New York, NY: Macmillan.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86. doi: 10.1207/s15326985ep4102_1

Knight, F. H. (1921). *Risk, uncertainty and profit*. New York, NY: Houghton Mifflin.

Kolmogorov, A. N. (1956). *Foundations of the theory of probability*. New York, NY: Chelsea. (Original work published 1933)

Korsgaard, M. T. (2019). Exploring the role of exemplarity in education: Two dimensions of the teacher's task. *Ethics and Education*, *14*(3), 271–284. doi: 10.1080/17449642.2019.1624466

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109. doi: 10.1037/a0020762

Kuntze, S., Aizikovitsh-Udi, E., & Clarke, D. (2017). Hybrid task design: connecting learning opportunities related to critical thinking and statistical thinking. *ZDM - International Journal on Mathematics Education*, *49*(6), 923–935. doi: 10.1007/s11858-017-0874-4

Lee, H. S., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal*, *9*(1), 68–96.

Lerman, S. (2000). The social turn in mathematics education. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 19–44). Westport, CT: Ablex.

Lester, F. K. (2005). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. *Zentralblatt für Didaktik der Mathematik*, *37*, 457–467.

Levinson, R., Kent, P., Pratt, D., Kapadia, R., & Yogui, C. (2012). Risk-based decision making in a scientific issue: A study of teachers discussing a dilemma through a microworld. *Science Education*, *96*(2), 212–233. doi: 10.1002/sce.21003

Lijnse, P. L. (1995). "developmental research" as a way to an empirically based "didactical structure" of science. *Science Education*, *29*(2), 189–199. doi: 10.1002/sce.3730790205

Lijnse, P. L., & Klaassen, K. (2004). Didactical structures as an outcome of research on teaching-learning sequences? *International Journal of Science Education*, *26*(5), 537–554. doi: 10.1080/09500690310001614753

Löfstedt, R. E. (1996). The Barsebäck nuclear plant case. *Energy Policy*, *24*(8), 689–696. doi: 10.1016/0301-4215(96)00042-0

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, *13*(1-2), 1–4. doi: 10.1080/10986065.2011.538291

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.

Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. B. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261–294). Cham, the Netherlands: Springer. doi: 10.1007/978-3-319-66195-7_8

Manuschevich, D. (2020, March 11). El gobierno mide mal lo que no le conviene, como las que marcharon en el 8m [The government wrongfully measures things it finds inconvenient, like those who marched on 8m]. *Bío Bío Chile*. Retrieved from `https://www.biobiochile.cl/noticias/opinion/tu-voz/2020/03/11/el-gobierno-mide-mal-lo-que-no-le-conviene-como-las-que-marcharon-en-el-8m.shtml`

Mardones, C. (2016, December 26). Liceos emblemáticos salen del ranking de los 50 mejores colegios en la PSU [Emblematic secondary schools get out of the 50 best schools in PSU ranking]. *La Tercera*. Retrieved from `https://www.latercera.com/noticia/liceos-emblematicos-salen-del-ranking-los-50-mejores-colegios-la-psu`

Mateos-Aparicio, G. (2002). Historia de la probabilidad (desde sus orígenes a Laplace) y su relación con la historia de la teoría de la decisión [The history of probability (from its origins to Laplace) and its relation to the history of Decision Theory]. In *Historia de la probabilidad y la estadística* (pp. 1–18). Madrid, Spain: Asociación de Historia de la Estadística y de la Probabilidad de España & AC.

Mateos-Aparicio, G. (2004). Evolución histórica de los métodos de decisión a partir de Laplace [Historical evolution of decision methods since Laplace]. In *Historia de la probabilidad y la estadística (II)* (pp. 139–155). Madrid, Spain: Asociación de Historia de la Estadística y de la Probabilidad de España & Delta.

Meder, B., & Gigerenzer, G. (2014). Statistical thinking: No one left behind. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 127–148). Dordrecht, the Netherlands: Springer. doi: 10.1007/978-94-007-7155-0_8

Mills, C. W. (2000). *The sociological imagination*. Oxford, UK: Oxford University Press. (Original work published 1959)

Ministerio de Educación. (2015a). *Matemática, Programa de Estudio, Cuarto Medio* [Mathematics, Study Programme, Fourth Middle]. Santiago, Chile: Unidad de Currículum y Evaluación, Ministerio de Educación de Chile. Retrieved from `http://www.curriculumenlineamineduc.cl/605/articles-34362_programa.pdf`

Ministerio de Educación. (2015b). *Matemática, Programa de Estudio, Tercero Medio* [Mathematics, Study Programme, Third Middle]. Santiago, Chile: Unidad de Currículum y Evaluación, Ministerio de Educación de Chile. Retrieved from `http://www.curriculumenlineamineduc.cl/605/articles-34361_programa.pdf`

Ministerio de Educación. (2017). *Propuesta Curricular para 3º y 4º Medio: Documento de Consulta Pública* [Curricular proposal for 3rd and 4th-middle: Public consultation document]. Santiago, Chile: Unidad de Currículum y Evaluación, Ministerio de Educación de Chile.

Ministerio de Educación. (2019). *Bases Curriculares: 3º y 4º Medio* [Curricular bases: 3rd and 4th-middle]. Santiago, Chile: Unidad de Currículum y Evaluación, Ministerio de Educación de Chile. Retrieved from `https://curriculumnacional.mineduc.cl/614/articles-91414_bases.pdf`

Monteiro, C., & Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education*(3), 187–207.

Mooney, E. S., Langrall, C. W., & Hertel, J. T. (2014). A practitional perspective on probabilistic thinking models and frameworks. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 495–507). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-007-7155-0_27

Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association*, *93*(444), 1253–1259.

Moore, D. S. (2010). *The basic practice of statistics* (5th ed.). New York, NY: Freeman.

Moore, D. S., & Cobb, G. W. (2000). Statistics and mathematics: Tension and cooperation. *The American Mathematical Monthly*, *107*(7), 615–630.

Murphy, P., Lunn, S., & Jones, H. (2006). The impact of authentic learning on students' engagement with physics. *The Curriculum Journal*, *17*(3), 229-246. doi: 10.1080/09585170600909688

National Committee on Mathematical Requirements. (1923). *The reorganization of mathematics in secondary education*. Boston, MA: The Mathematical Association of America.

Negt, O. (1975). *Sociologisk fantasi og eksemplarisk indlæring* [Sociological imagination and exemplary learning]. Roskilde, Denmark: Roskilde Universitetsforlag.

Nilsson, P. (2007). Different ways in which students handle chance encounters in the explorative setting of a dice game. *Educational Studies in Mathematics*, *66*(3), 293–315. doi: 10.1007/s10649-006-9062-0

Nilsson, P. (2014). Experimentation in probability teaching and learning. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 509–532). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-007-7155-0_28

Nilsson, P., Schindler, M., & Bakker, A. (2018). The nature and use of theories in statistics education. In D. Ben-Zvi, K. Makar, & J. B. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 359–386). Cham, The Netherlands: Springer. doi: 10.1007/978-3-319-66195-7_11

Niss, M. (1993). Centrale problemstillinger i matematikkens didaktik i 1990'erne [Central issues in mathematics education in the 1990s]. In E. S. Andersen (Ed.), *15. Nordiske LMFK-kongres* (pp. 93–116). Aarhus, Denmark: LMFK.

Niss, M. (1994). Mathematics in society. In R. Biehler, R. W. Scholz, R. Sträßer, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 367–378). Dordrecht, the Netherlands: Kluwer Academic.

Niss, M. (1996). Goals of mathematics teaching. In A. Bishop, M. K. Clements, C. Keitel-Kreidt, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 11–47). Dordrecht, the Netherlands: Kluwer Academic.

Niss, M. (2007a). The concept and role of theory in mathematics education. In C. Bergsten, B. Grevholm, H. S. Måsøval, & F. Rønning (Eds.), *Relating practice and research in mathematics education. Proceedings of Norma05, Fourth Nordic Conference on Mathematics Education* (pp. 97–110). Trondheim, Norway: Tapir.

Niss, M. (2007b). Reflections on the state of and trends in research on mathematics teaching and learning: From here to utopia. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1293–1312). Charlotte, NC: Information Age.

Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling*. Milton Park, UK: Routledge.

Nobre, S. (1989). The ethnomathematics of the most popular lottery in Brazil: The "Animal Lottery". In P. Keitel C. Damerow, A. Bishop, & P. Gerdes (Eds.), *Mathematics, education, and society* (pp. 175–177). Paris, France: UNESCO.

Pais, A., Fernandes, E., Matos, J. F., & Alves, A. S. (2012). Recovering the meaning of "critique" in critical mathematics education. *For the Learning of Mathematics*, *32*(1), 28–33.

Palm, T. (2009). Theory of authentic task situations. In L. Verschaffel, B. Greer, W. Van Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds: Modelling verbal descriptions of situations* (pp. 3–19). Rotterdam, The Netherlands: Sense.

Penteado, M. G. (2001). Computer-based learning environments: Risks and uncertainties for teachers. *Ways of Knowing*, *1*(2), 23–35.

People v. Collins, 68 Cal. 619, (Cal. 1968)

Petocz, P., Reid, A., & Gal, I. (2018). Statistics education research. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 71–99). Cham, the Netherlands: Springer. doi: 10.1007/978-3-319-66195-7_3

Pfannkuch, M. (2018). Reimagining curriculum approaches. In D. Ben-Zvi, K. Makar, & J. B. Garfield (Eds.), *International handbook of research in statistics education* (pp. 387–413). Cham, the Netherlands: Springer International. doi: 10.1007/978-3-319-66195-7_12

Piaget, J. (1970). *Genetic epistemology*. New York, NY: Columbia University Press.

Popkewitz, T. S. (2004). The alchemy of the mathematics curriculum: Inscriptions and the fabrication of the child. *American Educational Research Journal*, *41*(1), 3–34.

Popkewitz, T. S., Diaz, J., & Kirchgasler, C. (Eds.). (2017). *A political sociology of educational knowledge*. New York, NY: Routledge. doi: 10.4324/9781315528533

Popkewitz, T. S., & Lindblad, S. (2018). Statistics reasoning, governing education and making differences as kinds of people. In S. Lindbald, D. Pettersson, & T. S. Popkewitz (Eds.), *Education by the numbers and the making of society: The expertise of international assessments* (pp. 203–221). New York, NY & London, UK: Routledge.

Porter, T. M. (1986). *The rise of statistical thinking 1820-1900*. Princeton, NJ: Princeton University Press.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011). Role of context in risk-based reasoning. *Mathematical Thinking and Learning*, *13*(4), 322–345. doi: 10.1080/10986065.2011.608346

Prediger, S. (2008). Do you want me to do it with probability or with my normal thinking? Horizontal and vertical views on the formation of stochastic conceptions. *International Electronic Journal of Mathematics Education*, *3*(3), 126–154.

Prediger, S., Bikner-Ahsbahs, A., & Arzarello, F. (2008). Networking strategies and methods for connecting theoretical approaches: First steps towards a conceptual framework. *ZDM - Mathematics Education*, *40*(2), 165–178.

Prediger, S., & Rolka, K. (2009). Using betting games for initiating conceptual change in mathematics classrooms. *Asian Journal of Educational Research and Synergy*, *1*(1), 57.

Radakovic, N. (2015a). Pedagogy of risk: Why and how should we teach risk in high school math classes? *The Mathematics Enthusiast*, *12*(1,2&3), 307–329.

Radakovic, N. (2015b). "People can go against the government": Risk-based decision making and high school students' concepts of society. *Canadian Journal of Science, Mathematics and Technology Education*, *15*(3), 276–288. doi: 10.1080/14926156.2015.1062938

Radmehr, F., & Drake, M. (2019). Revised Bloom's taxonomy and major theories and frameworks that influence the teaching, learning, and assessment of mathematics: a comparison. *International Journal of Mathematical Education in Science and Technology*, *50*(6), 895–920. doi: 10.1080/0020739X.2018.1549336

Rosenthal, J. S. (2005). *Struck by lightning: The curious world of probabilities*. Canada: HarperCollins.

Rosenthal, J. S. (2015). Probability, justice, and the risk of wrongful conviction. *The Mathematics Enthusiast*, *12*(1,2&3), 11–18.

Ríos, S. (1998). La Teoría de la Decisión de Pascal a Von Neumann [Decision Theory from Pascal to Von Neumann]. In *Conferencias sobre Historia de la Matemática en el siglo XX, febrero-marzo de 1996.* Madrid, Spain: Real Academia de Ciencias Exactas, Físicas y Naturales.

Samuels, M. L., Witmer, J. A., & Schaffner, A. A. (2016). *Statistics for the life sciences* (5th ed.). Harlow, UK: Pearson. (Global Edition)

Sánchez Aguilar, M., & Blomhøj, M. (2010). The role of mathematics in politics as an issue for mathematics teaching. *Philosophy of Mathematics Education Journal*, *25*, 143–165.

Sánchez Aguilar, M., & Blomhøj, M. (2016). The role of mathematics in politics as an issue for mathematics teaching. In P. Ernest, B. Sriraman, & N. Ernest (Eds.), *Critical mathematics education: Theory, praxis and reality* (pp. 253–271). Charlotte, NC: Information Age.

Savard, A. (2015). Making decisions about gambling: The influence of risk on children's arguments. *The Mathematics Enthusiast*, *12*(1,2&3), 226–245.

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English language teaching*, *5*(9), 9–16. doi: 10.5539/elt.v5n9p9

Serrano, L., Batanero, C., Ortíz, J. J., & Cañizares, M. J. (1998). Heurísticas y sesgos en el razonamiento probabilístico de los estudiantes de secundaria [Heuristics and biases in secondary students' probabilistic reasoning]. *Educación Matemática*, *10*(1), 7–25. doi: 10.1017/CBO9781107415324.004

Shafer, G. (1996). The significance of Jacob Bernoulli's Ars conjectandi for the philosophy of probability today. *Journal of Econometrics*, *75*, 15–32.

Sharma, S. (2016). Probability from a socio-cultural perspective. *Statistics Education Research Journal*, *15*(2), 126–144.

Sharma, S. (2017). Definitions and models of statistical literacy: A literature review. *Open Review of Educational Research*, *4*(1), 118–133. doi: 10.1080/23265507.2017.1354313

Sharma, S., Doyle, P., Shandil, V., & Talakia'atu, S. (2011). Developing statistical literacy with year 9 students. *Set: Research Information for Teachers*(1), 43–50. (online)

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Silver, E. A., & Herbst, P. (2007). Theory in Mathematics Education Scholarship. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 39–67). New York, NY: NCTM - IAP. doi: 10.4065/81.11.1423

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivistic perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145. doi: 10.2307/

749205

Skovsmose, O. (1992). Democratic competence and reflective knowing in mathematics. *For the Learning of Mathematics*, *12*(2), 2–11.

Skovsmose, O. (1994). *Towards a philosophy of critical mathematics education*. Dordrecht: Kluwer.

Skovsmose, O. (1998). Linking mathematics education and democracy: Citizenship, mathematical archaeology, mathemacy and deliberative interaction. *Zentralblatt für Didaktik der Mathematik*, *30*(6), 195–203. doi: 10.1093/humrep/deg124

Skovsmose, O. (2005a). Foregrounds and politics of learning obstacles. *For the Learning of Mathematics*, *25*, 4–10.

Skovsmose, O. (2005b). *Travelling through education: Uncertainty, mathematics, responsibility*. Rotterdam, the Netherlands: Sense.

Skovsmose, O. (2008). Critique as uncertainty. In M. Menghini, F. Furinghetti, L. Giacardi, & F. Arzarello (Eds.), *The first century of the international commission on mathematical instruction (1908-2008).* Rome, Italy: Instituto della Enciclopedia Italiana.

Skovsmose, O. (2011). *An invitation to critical mathematics education*. Rotterdam, the Netherlands: Sense.

Skovsmose, O. (2014a). Critical mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 116–120). Dordrecht, the Netherlands: Springer.

Skovsmose, O. (2014b). *Critique as uncertainty*. Charlotte, NC: Information Age.

Skovsmose, O., & Borba, M. (2004). Research methodology and critical mathematics education. In P. Valero & R. Zevenbergen (Eds.), *Researching the socio-political dimensions of mathematics education: Issues of power in theory and methodology* (pp. 207–226). Dordrecht, The Netherlands: Kluwer.

Skovsmose, O., & Valero, P. (2002). Democratic access to powerful mathematical ideas. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 383–407). Mahwah, NJ: Lawrence Erlbaum.

Skovsmose, O., & Valero, P. (2012). Rompimiento de la neutralidad política: el compromiso crítico de la educación matemática con la democracia. In P. Valero & O. Skovsmose (Eds.), *Educación matemática crítica. una visión sociopolítica del aprendizaje y la enseñanza de las matemáticas.* (pp. 1–23). Bogotá, Colombia: Una empresa docente.

Spiegelhalter, D. (2014). Comments on probabilistic thinking. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (p. Back Cover). Dordrecht, the Netherlands: Springer.

Squires, J. (2005). Is Mainstreaming Transformative? Theorizing Mainstreaming in the Context of Diversity and Deliberation. *Social Politics: International Studies in Gender, State & Society*, *12*(3), 366–388. doi: 10.1093/sp/jxi020

Sriraman, B. (2007). On the origins of social justice: Darwin, freire, marx and vivekananda. *International Perspectives on Social Justice in Mathematics Education. The Montana Mathematics Enthusiast, Monograph*, *1*, 1–6.

Sriraman, B., & Lee, K. (2014). Commentary on perspective I: The humanistic dimensions of probability. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 117–119). Dordrecht, the Netherlands: Springer. doi: 10.1007/978-94-007-7155-0_6

Sriraman, B., & Nardi, E. (2013). Theories in mathematics education: Some developments and ways forward. In M. A. K. Clements, A. J. Bishop, J. Kilpatrick, C. Keitel, & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 303–325). New York, NY: Springer. doi: 10.1007/978-1-4614-4684-2

Steiner, H. G. (1985). Theory of mathematics education (TME): An introduction. *For the Learning of Mathematics*, *5*(2).

Taleb, N. N. (2001). *Fooled by randomness: The hidden role of chance in life and in the markets*. New York, NY: Random House.

Tele 13. (2017, December 26). PSU 2017: 3 de cada 10 alumnos no logra el puntaje mínimo para postular [PSU 2017: 3 out of 10 students do not reach the minimum score to apply]. *Tele 13*. Retrieved from `https://www.t13.cl/noticia/nacional/psu-2017-3-cada-10-alumnos-no-logra-puntaje-minimo-postular`

Televisión Nacional de Chile. (2018, October 31). Informe Especial: "Quintero, el plan que no funcionó" [Special Report: "Quintero, the plan that did not work"]. *24 Horas*. Retrieved from `https://www.24horas.cl/programas/informeespecial/informe-especial-quintero-el-plan-que-no-funciono-2849138`

Teppo, A. R. (2015). Grounded theory methods. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 3–21). Dordrecht, the Netherlands: Springer. doi: 10.1007/978-94-017-9181-6_1

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, *11*(3), 167–187. doi: 10.1007/BF01044641

Triantafillou, C., Psycharis, G., Bakogianni, D., & Potari, D. (2018). Enactment of inquiry-based mathematics teaching and learning: The case of statistical estimation. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 291–298). Umeå, Sweden: PME.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Vahey, P., Enyedy, N., & Gifford, B. (2000). Learning probability through the use of a collaborative, inquiry-based simulation environment. *Journal of Interactive Learning Research*, *11*(1), 51–84.

Valenzuela, J. P., Bellei, C., & de los Ríos, D. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, *29*(2), 217-241. doi: 10.1080/02680939.2013.806995

Valero, P. (1999). Deliberative mathematics education for social democratization in Latin America. *Zentralblatt für Didaktik der Mathematik*, *31*(1), 20–26. doi: 10.1007/s11858-999-0004-z

Valero, P. (2010). Mathematics education as a network of social practices. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education (CERME 6, January 28 – February 1, 2009)* (pp. LIV–LXXX). Lyon, France: Institut National de Recherche Pédagogique and ERME.

Valero, P. (2012). La educación matemática como una red de prácticas sociales [Mathematics education as a network of social practices]. In P. Valero & O. Skovsmose (Eds.), *Educación matemática crítica. una visión sociopolítica del aprendizaje y la enseñanza de las matemáticas* (pp. 299–326). Bogotá, Colombia: Una empresa docente.

Valero, P. (2017). Mathematics for all, economic growth, and the making of the citizen-worker. In T. Popkewitz, J. Diaz, & C. Kirchgasler (Eds.), *A political sociology of educational knowledge: Studies of exclusions and difference* (pp. 117–132). New York, NY: Routledge. doi: 10.4324/9781315528533-8

Valero, P., Andrade-Molina, M., & Montecino, A. (2015). Lo político en la educación matemática: De la educación matemática crítica a la política cultural de la educación matemática [The political in mathematics education: From critical mathematics education to the cultural politics of mathematics education]. *Revista Latinoamericana de Investigación en Matemática Educativa*, *18*(3), 7–20. doi: 10.12802/relime.13.1830

van Leeuwen, T. (2007). Legitimation in discourse and communication. *Discourse & Communication*, *1*(1), 91–112. doi: 10.1177/1750481307071986

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: Expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, *53*(2), 113–138.

Van Dooren, W., De Bock, D., Janssens, D., & Verschaffel, L. (2007). Pupils' over-reliance on linearity: A scholastic effect? *British Journal of Educational Psychology*, *77*(2), 307–321. doi: 10.1348/000709906X115967

Vergara Gómez, A., & Espinoza Ramírez, L. (2018). La matematización de los procesos de toma de decisiones; una mirada socioepistemológica [The mathematisation of decision making processes; a socioepistemological look]. In L. A. Serna & D. Pagés (Eds.), *Acta latinoamericana de matemática educativa* (Vol. 31, p. 599-606). México D.F, México: Comité Latinoamericano de Matemática Educativa A.C.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wagenschein, M. (1956). Zum Begriff des exemplarischen Lehrens [On the concept of exemplary teaching]. *Zeitschrift für Pädagogik*, *2*, 129–156.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46.

Wedege, T. (2006). Hvorfor stave problematik med q? - hvad, hvordan og hvorfor i matematikkens didaktik [Why to spell problematic with q? - what, how and why in didactics of mathematics]. In O. Skovsmose & M. Blomhøj (Eds.), *Kunne det tænkes? - om matematiklæring* (pp. 312–332). Copenhagen, Denmark: Malling Beck.

Wedege, T. (2016). Connecting the notion of foreground in critical mathematics education with the theory of habitus. In P. Ernest, B. Sriraman, & N. Ernest (Eds.), *Critical mathematics education: Theory, praxis and reality* (pp. 147–157). Charlotte, NC: Information Age.

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, *96*(1), 33–47. doi: 10.1007/s10649-017-9764-5

White, P., & Gorard, S. (2017). Against inferential statistics: How and why current statistics teaching gets it wrong. *Statistics Education Research Journal*, *16*(1), 55–65.

Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal*, *5*(2), 10–26.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. doi: 10.1111/j.1751-5823.1999.tb00442.x

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What is statistics? In D. Ben-Zvi, K. Makar, & J. B. Garfield (Eds.), *International handbook of research in statistics education* (pp. 5–36). Cham, the Netherlands: Springer International. doi: 10.1007/978-3-319-66195-7_1

Yin, R. K. (2009). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Zagzebski, L. (2010). Exemplarist virtue theory. *Metaphilosophy*, *41*(1–2), 41–57. doi: 10.1111/meta.2010.41.issue-1-2

Zagzebski, L. (2013). Moral exemplars in theory and practice. *Theory and Research in Education*, *11*(2), 193–206. doi: 10.1177/1477878513485177

Zieffler, A., Garfield, J. B., Delmas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40–58.

Appendix A

Interview Guide

Introduction to the interview

This is not a theme itself, but an introduction to the interview. It is a broad reminder of the experience. For this is necessary to make explicit:

- The purposes of the research project, namely the general *problématique* of coherence between justifications and teaching practice. A version was already stated in the proposal for classroom interventions in 2018.
- The purposes of the interview. In general, it consists of discussing the general *problématique* in light of the teaching experience, and therefore anchoring it into the classroom context. In particular, we are interested in both the actual experience and its potentialities.
- The form of the interview and collection of data: semi-structured and audio recording.
- To set the scene towards the first theme, we start with the face-value description of the sequence.

Theme: Students' work

The first main theme corresponds to the actual intervention that took place. Specifically, the interest is to remember the sequence in terms of students' work on it, and to validate part of the didactical analysis already made.

Sub-theme: Transformation/Impact on students

This sub-theme is a broad impression of the students.

- **Question:** What did the students say about the experience afterwards?
- **Prompt:** [1] –

[1] Non-specified prompts are those which were either not needed or came up as follow-up during the development of the interview.

Sub-theme: Engagement

It is relevant to know to what degree the students were engaged in the activity as a whole and what can be the **triggers** of such engagement.

- **Question:** Were students engaged in the activity, compared to their usual attitude in the mathematics classroom?
- **Prompt:** Students seemed to have fun.
- **Question:** To what extent was engagement due to the **context** (e.g. students' demonstration) of the task?
- **Prompt:**
- **Question:** To what extent was engagement due to the **mathematical content** (e.g. sampling and confidence intervals) of the task?
- **Prompt:** Are they fond of mathematics?
- **Question:** To what extent was engagement due to the **form of work** (e.g. inquiry, group work) of the task?
- **Prompt:** Had they worked this way before?
- **Question:** To what extent was engagement due to the **Hawthorne effect** (e.g. presence of a researcher and being recorded)?
- **Prompt:** They seemed comfortable with the camera anyway.

Sub-theme: Didactical interpretation

As part of the methodological proposal, the teacher can help validating part of the didactical analysis.

- **Question:** What is your interpretation of this exchange?
- **Prompt:** Selected recordings and transcripts. Sometimes the teacher provides her own interpretation, sometimes I provide and ask for her view.

Theme: Potential and possibilities

The third theme looks into the future. The relevant data to collect here is the teachers' experience-based opinion on what is the potential to work in a similar way on their own future practice.

Sub-theme: Learning potential

Connected to the previous didactical analysis of what *actually happened* now we move to analyse what *could have happened*.

- **Question:** Is it possible that this exchange could happen?
- **Prompt:** –
- **Question:** Can you think of other possible contexts and inquiries that can engage into a similar type of classroom activity?
- **Prompt:** –

Sub-theme: Relevance

Coming back to the general *problématique*, we want the teacher's input on the possible connections between critical justifications and the teaching practice of probability and statistics in high school.

- **Question:** To what extent are high school probability and statistics relevant for students to be empowered critical citizens?
- **Prompt:** –
- **Question:** To what extent is the work with socially-relevant inquiries as means for learning probability and statistics notions in high school?
- **Prompt:** –

Sub-theme: Working potential

Beyond the possibility for students to work in this type of activities, it is also relevant to know whether this is also possible for the teacher.

- **Question:** Is there or can there be a collaborative for of work with colleagues to implement similar activities?
- **Prompt:** How many mathematics teachers work here?

Sub-theme: Curricular framework

- **Question:** How was the match between the curricular contents (e.g. sample distributions) and the task (e.g. crowd size estimation)?
- **Prompt:** Do you think the activity was too difficult for students?
- **Question:** Is there a conflict between the need to cover the curriculum and the time invested in an inquiry-based lesson?
- **Prompt:** Do you remember our first meeting, when we discussed how to carry out the lessons?

Theme: Collaboration between researcher and teacher

This theme refers to the teacher's experience in working together with the researcher. The focus should not be into a good/bad evaluation, since there is potential bias into claiming it was a good experience just because it was different or to avoid conflict during the interview.

Sub-theme: Pressure

- **Question:** Did you feel pressured from me to collaborate on this research?
- **Prompt:** How did you feel about the original invitation to participate on this research?
- **Question:** Did you feel pressured from the school leadership to collaborate on this research?
- **Prompt:** I remember talking to the school's headmaster during those days, she seemed interested in our work.

Sub-theme: Participation

- **Question:** What was your role in the design of the sequence? (From imagined to actual situation)
- **Prompt:** To show ideas from the initial draft to the actual sequences, and notes from meetings.

Sub-theme: Aim balance

- **Question:** To what extent our aims (research/teacher) differed or conflicted one another?
- **Prompt:** –