“You are sentenced to timely justice”

Reducing case backlog at Family & Youth Court of Law Noord-Holland.

L. de Greef
s3067521

UNIVERSITY OF TWENTE.

de Rechtspraak

"You are sentenced to timely justice"

Reducing case backlog at Family & Youth Court of Law Noord-Holland.

This thesis is submitted as partial fulfilment of the degree of

Bachelor of Science in *Industrial Engineering & Management*

Tuesday, 08 July 2025, Enschede

Internal supervision

I.M.W. Schrader MSc
prof. dr. ir. E.W. Hans

External supervision

mr. N. Saft-Jacobsz
mr. B.J. van Putten

University of Twente
Universiteit Twente

Drienerlolaan 5
7522 NB Enschede
Tel. 053 489 9111

Court of Law Noord-Holland
Rechtbank Noord-Holland

Jansstraat 81
2011 RW Haarlem
Tel. 088 361 6100

Author

L. de Greef (Leon)
s3067521
l.degreef-1@student.utwente.nl

Preface

“Wie niet twijfelt, leert niet” (Those who do not doubt, do not learn) is painted on a building opposite the court building at the Jansstraat in Haarlem. It is a proverb belonging to D.V. Coornhert (1522-1590). This versatile and talented humanist was known for his extreme work ethic. He would work and study from four in the morning till ten in the evening to keep up with all political, theological, and philosophical developments of his time. Although Coornhert was born over half a millennium ago, his spirit still lives on in the Jansstraat.

Even though the judges, clerks, and supporting staff of section Family & Youth do not share exactly the same work hours as Coornhert – at least I hope – they do share an admirable work ethic and an incredible commitment to deliver justice to the people that need it most.

Family and Youth covers cases ranging from divorce to youth delinquency, from supervisor orders to compulsory care, and everything in between relating to families, relations, and youth in vulnerable positions. Delivering justice in these cases requires a lot of training and experience due to the complex case contents and the immense impact of verdicts on justice seekers’ lives. It is important that verdicts are delivered with quality and in time.

At team Family & Youth there is a lot of “twijfel” about case contents. In the words of Coornhert, this is a welcome phenomenon. Judges and clerks consult each other about the approach to take, or the verdict to speak. This ensures the quality of the verdicts.

In spite of that, unfortunately, delivering justice in time is not as self-evident as delivering justice with quality. For it is the *raison d'être* of this thesis, the problem is investigated and solutions are proposed.

The establishment of this thesis is far from an individual undertaking. A variety of people have added value – directly and indirectly – to my capacity to create this report. I would like to name them here to show them my profound appreciation.

First of all, I would like to thank the teaching staff of the BMS faculty and the nonteaching staff of the UT for both their direct and indirect support in the creation of this thesis. These people opened my eyes in varying ways with professionality and enthusiasm to learn about myself, others, and the world.

I am grateful to the external supervision of Nancy and Bob at the Court of Law Noord-Holland. They have provided me with huge freedom to research what I wanted. Also, their willingness to help and provide me with information was almost without limits. Since I had no fixed workplace, I had the opportunity to sit with different Family & Youth staff weekly (more than twenty different workplaces), resulting in many conversations about perspectives on problems and experiences from the Family and Youth team. Therefore, I would like to thank the entire Family & Youth team for their enthusiasm and determination to help in whatever way possible.

This thesis would have not come this far without the internal supervision of Ieke and Erwin. They were always available for helpful, supportive, critical, and honest insights, feedback, and solutions.

Dankjulliewel!

Leon de Greef,
Enschede, 08 July 2025

Management summary

This research attempts to provide recommendations to reduce case backlog at the Family and Youth section of the Court of Law Noord-Holland. The core problem of the existence of case backlog is determined to be the presence of idle time in the hearing schedule – which is the *leitmotif* of this thesis. This is caused by the inflexibility of the hearing schedule that results in the inevitable scheduling of idle time when a buffer is needed.

Case backlog and corresponding waiting times should be reduced to a minimum as it has negative consequences for the work pressure of judges and clerks, and for the lives of justice seekers.

Based on the analysis of the current scheduling process and literature available concerning linear programming in the judicial system as well as in the broader service sector, we propose a Mixed Integer Linear Program (MILP). The MILP is considered because the solution should provide integer outcomes and because earlier done research successfully applies this technique in hearing scheduling in the judicial system. The created MILP minimises the idle time between hearings by adjusting the hearing length – the number of cases that are batched in a hearing.

We identify the queuing system of Family and Youth to be able to make predictions about case backlog and waiting time evolution under the found solutions from the MILP. Family and Youth's queue is found to be best represented by a $G/G/z$ queue. Here, the arrival rate of cases is *Negative Binomial* distributed, and the service rate is generally distributed.

With the available input data, the MILP is used to determine the optimal hearing length per case type in a sensitivity analysis. The found optimal hearing lengths are examined in a robustness analysis. Based on the known queuing parameters, estimations are made with the *Allen-Cunneen formula* on the backlog and waiting time evolution under varying buffer sizes.

Although we cannot make a definitive and workable recommendation concerning the optimal hearing length currently – due to a lack of sufficient and reliable case input, output, and backlog data – there are still valuable conclusions originating from this research:

First, we find that the backlog size and waiting time is higher, and judge utilisation lower, under the realised performance of the current hearing lengths compared to their theoretical performance. This shows the existence of idle time and shows why Family and Youth has developed case backlog, even though they have a large enough judge capacity in theory.

Second, the MILP is able to provide optimal hearing lengths – different from the current hearing lengths – performing better in terms of service rate, approximated backlog size, and waiting time.

Third, for the scientific community, this research examines the optimal length of hearings. Our literature search shows that this is a previously unexplored area within research in the judicial system.

Table of Contents

Table of Figures

Table of Tables

Abbreviations & Glossary

Dutch	Abbreviation	English
alimentatie	ALI	alimony
arrondissement		court district
bestuursrecht		administrative law
bodemprocedure		long stream
boek 1	BK1	book 1
civielrecht		civil law
echtscheiding	ES	divorce
echtscheiding verdeling	ESVD	divorce division
enkelvoudige kamer	EK	single-judge chamber
Familie & Jeugd	F&J / F&Y	Family & Youth
gezag & omgang	G&O	authority & intercourse
griffier		clerk
handelsrecht		commercial law
jeugdstrafrecht	JS	youth criminal law
kort geding	KG	summary proceeding
korte stroom	KS	short stream
meervoudige kamer	MK	multi-judge chamber
ondertoezichtstelling	OTS	supervision order
Openbaar Ministerie	OM	Public Prosecutions Office
raadkamer	RK	judge's chamber
rechtbank		court (of law)
rechter		judge
secretaris		clerk
strafrecht		criminal law
spoedprocedure	KS	short stream
verplichte zorg	VZ	compulsory care
vonnis		verdict
voorlopige voorziening	VOVO	interim measure
zaak		case
zitting		hearing

Table 1: Abbreviations and glossary (de Rechtspraak, 2025a).

1 Introduction

Chapter 1 provides a clarification of the problem to be solved in this thesis together with the solving process. *Section 1.1* gives an introduction about the organisation at hand – section Family & Youth at the Court of Law Noord-Holland – and its internal structure. *Section 1.2* searches for the core problem to solve, and *Section 1.3* proposes the research methodology and research questions.

1.1 Background

De Rechtspraak is an umbrella term for all organizations involved in the judiciary of the Netherlands. Their collective mission is that de Rechtspraak "protects rights and freedoms, upholds the democratic rule of law, safeguards the proper application of the law and ensures that decisions are handed down by independent, impartial, ethical and competent judges" (de Rechtspraak, 2021a).

As part of de Rechtspraak, the Court of Law Noord-Holland has jurisdiction over one of eleven court districts (arrondissementen) of the Netherlands (*Figure 1*). As district court, it is comprised of three courts: Alkmaar, Haarlem, and Zaanstad. As of 1 January 2025, the latter of the three is temporarily closed due to renovation works. Together they have jurisdiction over the province of Noord-Holland – excluding the Amsterdam region. The courts adjudicate matters in the areas of administrative law (bestuursrecht), civil law (civiel recht), and criminal law (strafrecht). Within the civil law branch, section Family & Youth (Familie & Jeugd) is located. Family & Youth is based both in Alkmaar and Haarlem, and speaks out verdicts for over ten thousand cases per annum. These are distributed over an extensive range of case types, which are structured in *Appendix A.1 F&Y structure* (de Rechtspraak, 2023, 2024b; Rechtbank Noord-Holland, 2025a).

Section Family & Youth at the Court of Law Noord-Holland is indicated with "F&Y" from here onwards.

Figure 1: Court district of Court of Law Noord-Holland (Rechtbank Noord-Holland, 2025b).

1.2 Problem identification

In recent years, several F&Y judges have retired without there being sufficient successors. This is part of a greater problem within the judicial system; fewer people devote their careers into becoming judges. Because of the emotionally and morally difficult case content, F&Y needs experienced judges to uphold the quality of justice delivered. These experienced judges are becoming increasingly scarce, causing the need for F&Y to organise its processes more efficient to keep up with incoming case demand (Chavannes, 2019; de Rechtspraak, 2019; Duijneveldt et al., 2016).

F&Y experiences problems with sizable and growing case backlog in long stream cases[1] – and subsequently also in short stream cases[2]. Short stream case demand grows because justice seekers need to have preliminary arrangements until the definitive settlement in their long stream case. Thus, the longer it takes for a long stream case to be settled, the more short stream cases are needed. Because of the intrinsic urgency of short stream cases, they have priority over long stream cases, causing the waiting time for all justice seekers to increase. This is contrary to Article 6 of the *European Convention of Human Rights*, which states that “everyone is entitled to a fair and public hearing within a reasonable time” (European Court of Human Rights, 2013).

The growing case backlog also results in increased workload for judges and clerks. Due to the additional growth in short stream cases, the associated judges also experience unevenly distributed workload because of rushing and standing still due to the short notification of the cases and corresponding short preparation times.

Long stream cases that need to wait longer to reach a settlement grow in size while on the shelf, due to additional information coming in or new events occurring concerning the case. Therefore, the longer a case remains in the system the more time is needed to finalise it.

This section identifies the action problem, develops a problem cluster, and determines the core problem to address.

1.2.1 Action problem

Based on the problem identification, we develop the action problem as follows:

> *The size of case backlog of F&Y should be reduced to reduce waiting times for justice seekers without increasing the judge and clerk capacity.*

We choose this action problem since case backlog is large – and growing – for several case types within F&Y. The reality is that the case backlog is too large; the size of the case backlog is determined in *Section 2.3.3*.

1.2.2 Problem cluster

Based on conducted unstructured interviews with judges, clerks, and the hearing scheduling team, we compose a problem cluster to grasp the relationships and hierarchy between F&Y’s problems. Unstructured interviews are chosen to not limit the answer options of the participants, to gain varying answers in different areas. Furthermore, available data files are used to verify and validate the severity and presence of observed problems. The problem cluster in *Figure 2* also indicates the sections where the problems are elaborated on.

---

[1] Long stream case (bodemprocedure): A normal and extensive procedure in court (de Rechtspraak, 2025c).

[2] Short stream case (korte stroom/spoedprocedure): A(n) (urgent) procedure that acts as a preliminary measure for a long stream case (de Rechtspraak, 2025c).

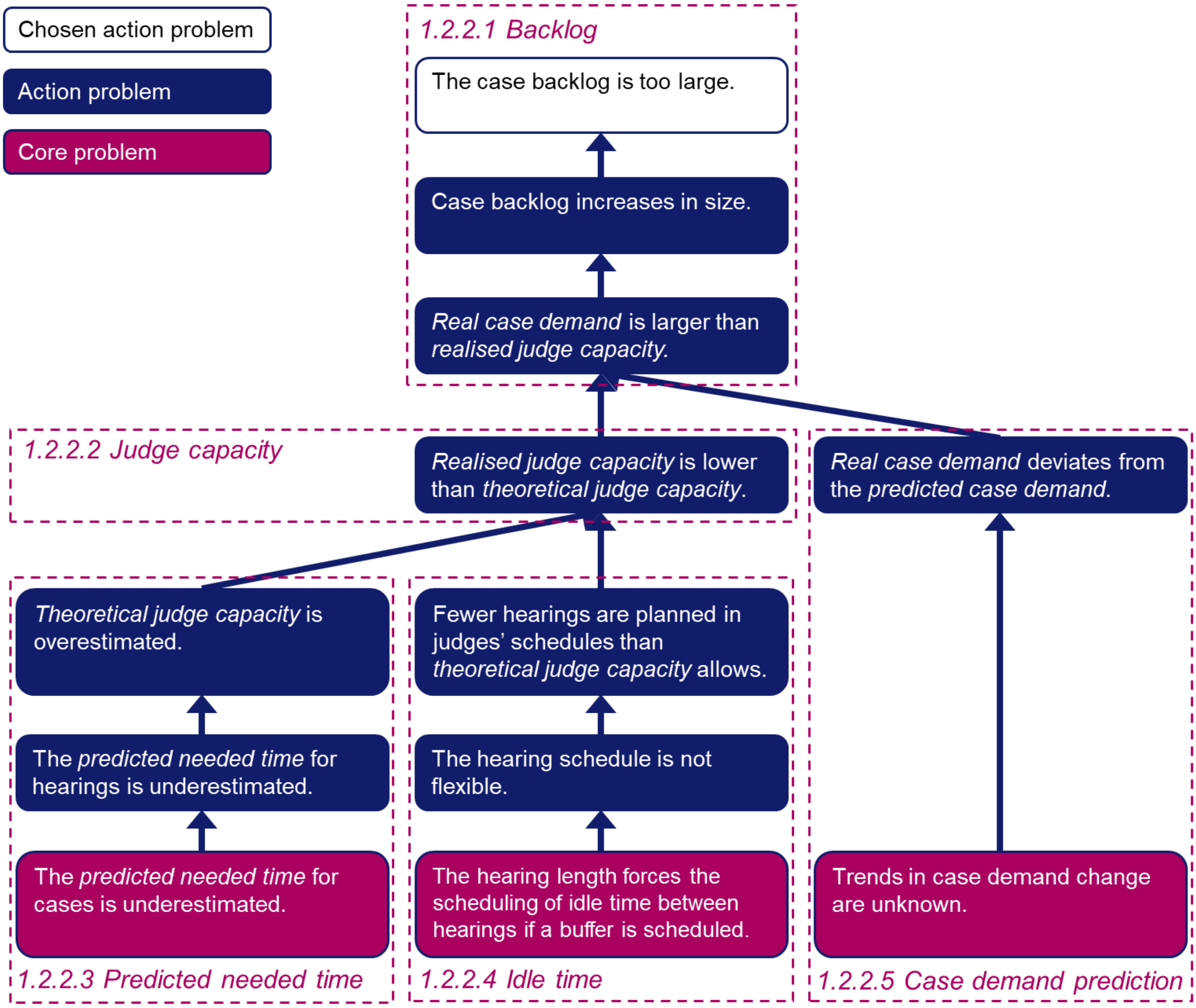


Figure 2: Problem cluster of F&Y including reading guide.

1.2.2.1 Backlog

We define backlog as "all cases present in the system." This is assumed because it is difficult to determine precisely for all case types when cases are in the system for "too long" (i.e. not all case types have strict deadlines).

The case backlog came to existence because the *real case demand* – the cases that actually need to be scheduled – is larger than the *realised judge capacity* – the number of cases that judges can process in a given timeframe.

The cause of this problem is twofold. First, the *realised judge capacity* is lower than the *theoretical judge capacity*, and second, the *real case demand* deviates from the *predicted case demand*.

1.2.2.2 Judge capacity

The utilisation of judges is lower than the *theoretical judge capacity* – the number of cases a judge can process in a given timeframe based on the contractually defined number of hours it has available for primary tasks. For example, *Judge A* has a contract of 36 hours per week. *Judge A* will not be able to spend all those hours on primary tasks (i.e. cases and hearings), because there are secondary tasks (e.g. meetings, courses et cetera) that also need to be dealt with. Therefore, a factor is defined in the contract that determines the number of hours a judge can work on primary tasks. If – for instance – the factor is 0.8 for *Judge A*, the number of hours that is available for cases and hearings in a week is approximately $36 * 0.8 \approx 29$. The weekly *theoretical judge capacity* of *Judge A* is the number of cases that *Judge A* can process in these 29 hours.

The *realised judge capacity* is lower than the *theoretical judge capacity* because the latter is overestimated. Moreover, fewer hearings are planned in the schedules of judges than possible according to the *theoretical judge capacity*. Therefore, the utilisation of judges is suboptimal.

1.2.2.3 Predicted needed time

The *theoretical judge capacity* is overestimated because the *predicted needed time* for hearings is systematically incorrect since the *predicted needed time* for cases is systematically incorrect. The needed time is underestimated; hence, causing the *theoretical judge capacity* to be overestimated, in turn causing the *realised judge capacity* to be always lower than the *theoretical judge capacity*.

Hearings have a predicted time that corresponds with the needed time to prepare and finish all cases within them. *Appendix A.2 Time prediction* provides an overview of the predicted needed times used by F&Y. To illustrate, *Judge A* has 29 hours to work on primary activities per week, this means that according to the *predicted needed time* for hearings, it can process $29\, primary\ hours / (4\, preparation\ hours + 4\, hearing\ hours + 4\, finishing\ hours) = 2.42$ OTS hearings per week – meaning $2.42\, hearings * 4\, cases\ per\ hearing = 9.68$ OTS cases. However, if in reality the needed time for a hearing is more than the *predicted needed time*, then the number of hearings and cases a judge can process in a timeframe is less. This results in an overestimated *theoretical judge capacity*.

1.2.2.4 Idle time

Less hearings are planned in judges' schedules because the hearing schedule does not offer enough flexibility. This because idle time needs to be scheduled between hearing processes when a buffer is needed. To clarify, in every hearing process a predetermined number of cases is planned. For instance, in an ES hearing process four cases, in an ESVD hearing process two cases. This rigid structure is present because two blocks of four hours are used for hearing scheduling. The morning hearing is from 9:00 till 13:00, and the afternoon hearing from 13:00 till 17:00 (*Figure 3*). The VZ case type is an exemption to this since it takes place outside the court and lasts an entire working day (9:00-17:00).

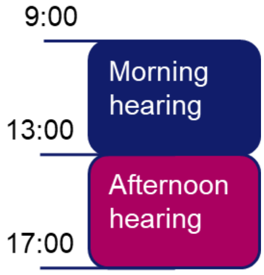


Figure 3: Visualisation of the morning and afternoon hearing blocks.

The need to schedule idle time arises with these hearing lengths – the number of cases planned in a hearing. For example, *Judge B* is assigned to an OTS hearing process. Four hours of preparation time, a hearing, and four hours of finishing time need to be scheduled. When starting on Monday, the preparation can be done on Monday morning, the hearing can be done in the Monday afternoon hearing block, and the finishing can be done on Tuesday morning. *Hearing process 1* is finished and another hearing process (*Hearing process 2*) can be scheduled; with preparation on Tuesday afternoon, hearing on Wednesday morning, and finishing on Wednesday afternoon et cetera. However, if the finishing of *Hearing process 1* takes longer than predicted, then there is not enough preparation time for *Hearing process 2*. A lack of preparation time is something that is never allowed to occur (de Rechtspraak, 2021b).

A buffer is scheduled between hearing processes to coop with the uncertainty in *real needed time*. However, an entire daypart (four hours) has to be scheduled between hearing processes irrespective of the length of the needed buffer (i.e. one, two, or three hours). This is caused by the hearing block length of four hours. The extra time that is scheduled on top of the buffer is the idle time. The size of the idle time is dependent on the hearing length – the number of cases in a hearing.

Figure 4 shows the disparity between the theoretical and realised schedule of the given example. The inevitability of having idle time in the schedule when a buffer – smaller than four hours – is needed results in unutilised judge capacity.

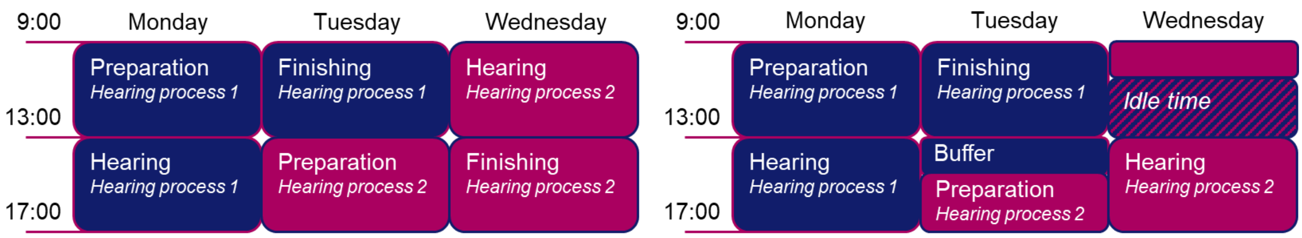


Figure 4: Ideal schedule (left) and realised schedule (right) for judges.

If the needed buffer is exactly four hours, there is no idle time, hence no problem. We assume that this not the case. *Section 2.2.4* justifies this assumption.

One could argue that the underestimated *predicted needed time* is the root cause of the need to schedule a buffer. However, a buffer can be hearing type dependent. Furthermore, if the *predicted needed time* is recalculated – and does not neatly correspond with the four hours blocks – the need to forcibly schedule idle time is still present due to the rigid hearing block structure. So, improving the *predicted needed time* does not solve the problem by itself.

1.2.2.5 Case demand prediction

The *real case demand* deviates from the *predicted case demand* because trends in case demand change are unknown. Every year, an estimation is made of the case demand in the upcoming year, based on one year of historical data. Therefore, it is unknown whether there is a trend in demand change. Furthermore, the reliability of a sample size of one year is arguably not convincing nor reliable to use as prediction for the upcoming year.

1.2.3 Core problem

Based on the problem cluster, we identify three core problems causing the chosen action problem and determine their addressability below.

1. *The predicted needed time for cases is underestimated.*

This first core problem is not successfully addressable within the time period that is available for the execution of this research (ten weeks). Moreover, predicting the needed time for cases and hearings requires both historical data and knowledge of judicial processes and case types. Neither the data, nor the knowledge is available. Based on interviews with judges, clerks, and F&Y's hearing scheduling team, we find several lines of reasoning on why the time prediction is incorrect. From the fact that some judges and clerks cannot "work as fast" as prescribed, to the reason that the time predictions were done in a time when not all cases were "complex and difficult" – a typical case of chronocentrism (Fowles, 1974).

2. *The hearing length forces the scheduling of idle time between hearings if a buffer is scheduled.*

This second core problem is addressable because this problem is concerned with the way the schedule is made instead of predicting and forecasting the needed time and demand, respectively. Here, unlike for *Core problem 1*, the necessary data (i.e. historical schedules, judge and clerk specialisations, primary hours), knowledge, and (mathematical) techniques that need to be implemented are available. Addressing the schedule in a proper way also makes sure that the effect of deviations in *real needed time* for cases and hearings is mitigated sufficiently. Furthermore, it brings flexibility in the schedule when we address this problem.

3. *Trends in case demand change are unknown.*

This third problem is not addressed because of the *relevance rule* (Heerkens & Winden, 2017). This problem is deemed possible, yet less beneficial, to solve than *Core problem 2*. For the reason that we assume that a good hearing schedule is able to maximise judge utilisation independent of the case demand. Furthermore, we expect that a solution to the hearing scheduling will have a larger impact on case backlog reduction than improving the case demand prediction.

1.3 Solution planning

Here, we propose and evaluate the chosen research design. Subsequently – based on the problem identification – we propose the main research question together with sub-questions and their solving approaches.

1.3.1 Research design

1.3.1.1 Methodology

The environment where the problem is located creates the necessity for a flexible methodology. The *Managerial Problem-Solving Method* (*MPSM*) is chosen since it is a versatile method which is adjustable to the problem at hand. The *MPSM* consists of seven phases: The first part (phase 1-3) is connected to finding the problem and looking into ways to solve the found problem; the second part (phase 4-6) is comprised of phases which deal with the creation and implementation of the solution; it is concluded with the evaluation of the solution (phase 7) (Heerkens & Winden, 2017).

Figure 5 shows the structure of the thesis plotted out together with the seven *MPSM* phases.

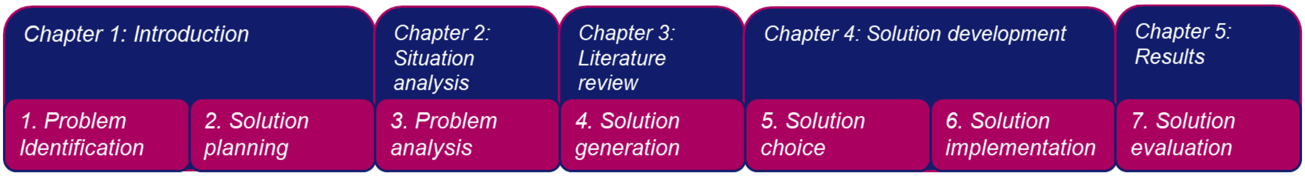


Figure 5: The distribution of MPSM phases within the thesis chapters.

Upon encountering a knowledge problem, one steps out of the *MPSM* and enters the *Research Cycle* (*Figure 6*). The *Research Cycle* is an eight phases cycle to solve the knowledge problem at hand. Upon completion, one steps back into the *MPSM* and continues. It is possible to enter the *Research Cycle* from every *MPSM* phase except for phase 5. The process is nonlinear, but circular. After phase 8, one goes back to phase 1 to reiterate the cycle until all knowledge questions in the respective phase are resolved.

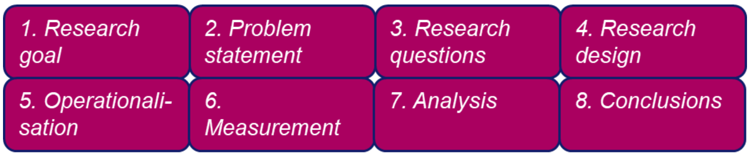


Figure 6: The eight steps of Heerkens & Winden's Research Cycle.

1.3.1.2 Research question

The research question is concerned with finding a fitting solution to the identified core problem:

> *The hearing length forces the scheduling of idle time between hearings if a buffer is scheduled.*

To reduce the case backlog, the idle time needs to be reduced. As *Section 1.2.2.4* describes, the size of the idle time is dependent on the hearing length – the number of cases in a hearing. Therefore, we propose the following research question:

> *How can section Family & Youth at the Court of Law Noord-Holland reduce case backlog by adjusting the number of cases in hearings?*

To answer the research question, we propose sub-questions per chapter to answer different aspects of the research question. Below, an overview is given of the sub-questions based on the phases of the *MPSM*. More specific questions are raised per chapter, together with the research design of the solving process.

1.3.1.3 Problem analysis

> *Chapter 2 Situation analysis – How does the hearing scheduling process work currently?*

This sub-question aims to build understanding of the current situation at F&Y to build knowledge about the hearing scheduling process.

We raise three questions to gain specific insights into the hearing scheduling, buffers and idle time, and cases of F&Y.

a. *How is the hearing schedule made?*

The answer to this question is visualised by means of a Business Process Model and Notation (BPMN). Necessary data is provided by conducting unstructured interviews with the hearing scheduling team of F&Y. The choice for unstructured interviews over (semi-) structured is that the data is gathered *in itinere* – in a timeframe of several weeks. The outcomes of the interviews are verified by asking multiple relevant stakeholders. Next to interviewing, a look is taken at the guidelines and rules within the court about the scheduling process to find restrictions, limitations, and constraints related to the scheduling process.

b. What is the current share of the buffer and idle time in the judges' schedules?

An answer to this question is essential since it represents the severity of the core problem at hand. The scheduling software of F&Y is used as a source of the needed secondary data. Furthermore, we provide estimations about the sizes of the buffer and idle time.

c. What is the case input, output, and backlog of F&Y?

We answer this question to provide a broader picture of the current situation and to inform ourselves about the reality of the chosen action problem to solve. Secondary data is obtained from the Court of Law Noord-Holland and from de Rechtspraak.

1.3.1.4 Solution generation

Chapter 3 Literature review – What literature exists for scheduling?

Based on the knowledge and techniques obtained during the study programme preceding the writing of this thesis – *Industrial Engineering & Management*, linear programming and queuing theory are known and familiar mathematical techniques which can be applied to the problem at hand (Chen et al., 2010; Meijer, 2022, 2023; Winston, 2022).

Linear programming is defined as the "efficient assigning of limited resources to the specified activities in order to maximize the interest and minimize the cost" (Babaei et al., 2015; Bixby, 2012). This technique – including its variations (i.e. (mixed) integer linear programming) – is applicable to the problem of F&Y since it is necessary to assign the limited resources (i.e. judges) to the activities (i.e. hearings) efficiently to minimize costs (i.e. idle time). With linear programming, a solution is sought for the core problem by minimising the idle time. The literature is needed to provide an understanding of existing linear programming applications as well as techniques to apply to the situation at hand.

Queuing theory is concerned with the "mathematical study of waiting lines" (Sundarapandian, 2009). The waiting lines – queues – of F&Y are to be used to determine the current state, and the effect of the solution. With queuing theory, the utilisation of judges and the waiting time of justice seekers is measured; two factors related to the chosen action problem to be solved. Furthermore, queuing theory is a tool to predict the increase and decrease of backlog. Literature is needed to build understanding about different queuing systems and the queuing system that best describes F&Y.

We conduct a literature review to broaden the horizon about the abovementioned mathematical techniques in scheduling. This chapter's sub-question is explored further by answering five questions:

a. Which linear programming techniques exist for scheduling in the judicial system?
b. Which linear programming techniques exist for scheduling in the service sector?
c. Which linear programming techniques exist for reducing makespan in production management?
d. Which queuing techniques exist for scheduling in the judicial system?
e. Which queuing techniques exist for scheduling in the service sector?

Since both linear programming and queuing theory in the judicial system are limited researched topics, we assume that the size of relevant literature is also limited. Therefore, questions (*b* & *e*) are raised where the scope of the literature search is broadened to find relevant literature in other service sectors – which share parallels with the judicial system – to be able to apply those researched insights to F&Y. We raise question *c* since the problem solving approach shows similarities with the process of reducing makespan by changing batch sizes in production management (İnce et al., 2024).

The aim of broadening the horizon about linear programming and queuing theory is twofold. First, conducted research is examined to see what has already been addressed in the literature in both the judicial sector as well as in the service sector in general; to find gaps in the literature to address. Second, knowledge needs to be grown by means of literature to find techniques which can help address F&Y's problem.

1.3.1.5 Solution choice & implementation

Chapter 4 Solution development – How will the solution for F&Y be realised?

In the solution creation, we incorporate the feasible, usable, and effective techniques found in the literature. These solutions are materialised by answering the two following questions.

a. *How does the hearing length influence the idle time length between hearings?*

To answer this question, we develop a linear programming model in which the objective is to minimise the forcibly scheduled idle time by adjusting the hearing length. As *Section 1.2.2.4* already explained, the solution will still be valid when the *predicted needed time* for cases and hearings, or the needed buffer size, is recalculated and improved. This because the solution will bring flexibility in the hearing schedule since it minimises idle time for judges.

b. *What are the parameters of F&Y's queuing system?*

With this question, the queuing system of F&Y is determined together with the corresponding parameters. This information is needed to determine how backlog and waiting times evolve when the solution is implemented. It provides a base for comparison between the current situation and the solution. This question is answered by looking into queuing theory to explain the relations between backlog, utilisation, and waiting times.

1.3.1.6 Solution evaluation

Chapter 5 Results – To what extent does the solution contribute to the reduction of case backlog?

In *Chapter 5*, we conduct experiments to examine the results of the linear programming model. A sensitivity and robustness analysis are performed to find the optimal hearing lengths under different buffer lengths, and to examine the development of idle time under those hearing lengths, respectively. These experiments are needed because the real needed buffer size is unknown.

Furthermore, the waiting times for justice seekers, the utilisation of judges and the backlog evolution is discussed under the different parameter values by means of a queuing analysis.

1.3.2 Scope

The scope of this report is concerned with the hearing scheduling of the Family & Youth section of Court at Law Noord-Holland, with an additional focus on the Haarlem court.

The weight of the solution lies in the improvement of the judges' schedule, since they are the scarcest resource of F&Y – which *Section 2.2.4* shows. Although the solution can yield optimisation of the judges' schedule, it can be detrimental for the utilisation of the clerks.

We attain the borders between the current hearing blocks (*Figure 3*). This to fit in the existing separation of the morning and afternoon block as it is the *modus operandi* of Court of Law Noord-Holland; and to not have a hearing spread over two days.

(Seasonal) trends in changing case demand are left out of the scope since there is insufficient data to prove the existence of these. In the case of existing seasonal trends in case demand, it can be that the solution will not provide an optimal solution.

2 Situation analysis

Chapter 2 describes and explores the current situation at Family & Youth of Court of Law Noord-Holland by answering this chapter's question – *How does the hearing scheduling process work currently?* We answer the sub-questions in their respective sections. *Section 2.1* addresses the question *How is the hearing schedule made? Section 2.2* answers the question *What is the current share of the buffer and idle time in the judges' schedules?* And *Section 2.3* is concerned with answering *What is the case input, output, and backlog of F&Y?*

2.1 Current hearing scheduling process

In this section, the hearing scheduling process is explored, together with the related constraints. Furthermore, the differences between case types and hearing types is explained.

2.1.1 Hearing scheduling process

A *Business Process Model and Notation* – BPMN for short – "creates a standardized bridge for the gap between the business process design and process implementation" according to its creator, the Object Management Group (2010). BPMNs visualise processes to improve comprehension of the process concerned.

With the acquired knowledge from the unstructured interviews with the hearing scheduling team, we create a BPMN to visualise the hearing scheduling process. *Appendix B.1 BPMN* provides the full *BPMN*, *Figure 7* shows a simplified version.

The scheduling process starts at the hearing scheduler, who – based on the budget – determines the number of hearings to schedule per case type (the case mix). In reality, these budget predictions are not materialised due to the *realised judge capacity* – which is smaller than the *predicted hearing demand*. The hearing scheduler makes the hearing schedule per month, hence adjusts the annual prediction numbers to that. When there is a hearing to schedule, the hearing scheduler picks a hearing block in the schedule to plan it; otherwise, another case type is chosen, a new month is chosen, or the process is ended. If the planner finds a suitable and available judge and clerk, the hearing is scheduled in the system. The cases treated in a hearing are planned by the case planner. The courtroom where the hearing takes place is scheduled by the courtroom planner.

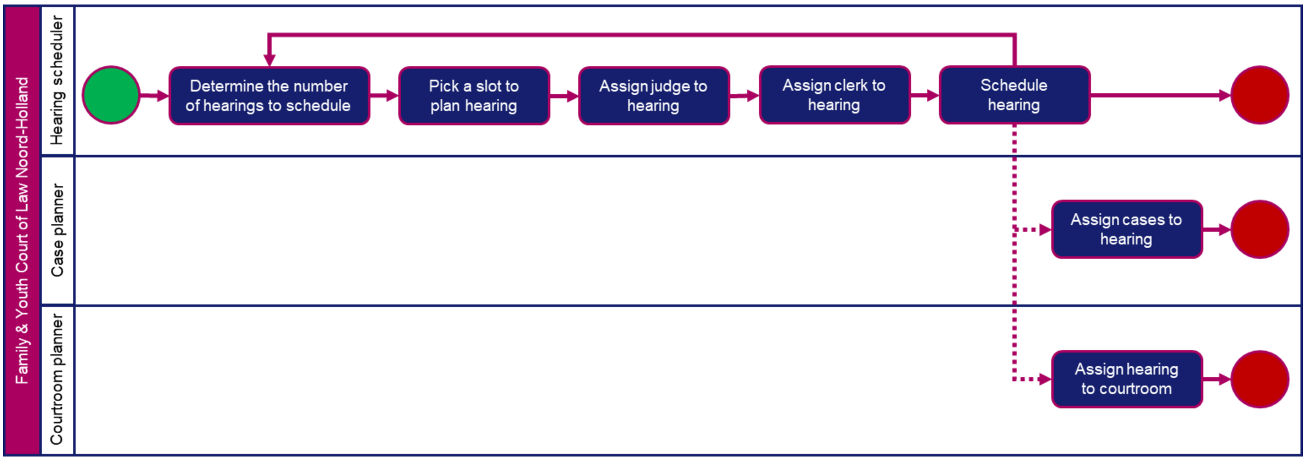


Figure 7: Simplified BPMN of the scheduling process of F&Y.

The number of judges, clerks, and courtrooms is capped in reality. For reasons of simplicity and clarity, this is left out of the BPMN in *Appendix B.1 BPMN*. Also, the case scheduling of the case planner and the courtroom scheduling of the courtroom planner are simplified since it is out of the scope of this research.

The BPMN captures a single-judge hearing. For multi-judge hearings, the process remains similar except for the fact that the search for a judge has to be executed threefold.

2.1.2 Scheduling constraints

In the hearing scheduling process, the hearing schedulers need to take into account constraints. We elaborate on the most obvious and impactful constraints below.

2.1.2.1 Hearing order

Legally, all hearing types can be scheduled after each other independent of the severity or length of the preceding hearing type. Besides personal preferences of judges and clerks (e.g. no short stream hearings after each other), the hearing order is not dependent on the hearing types that need to be scheduled.

2.1.2.2 Holidays

The length of a judge’s holiday has an influence on which hearing types they can work in the week before and after their holiday. This is because – depending on the holiday length – the verdict deadline can be surpassed (de Rechtspraak, 2025b). Also, not all hearing types can be performed in the week after the holiday because there would be too little preparation time before the hearing.

In *Table 2*, the opaque coloured cells indicate that it is not possible to start a hearing process in the week before a judge’s holiday. The striped coloured cells indicate a possibility under a condition. For the KS hearing type, the condition is that a hearing can only be performed on the Monday before the holiday if the judge works fulltime (independent of the length of holiday). The FAM hearing type can only be performed on the Monday before the holiday – also independent of the holiday length.

Hearing type	**Length of holiday (weeks)**		
	1	**2**	**>2**
KS			
G&O			
FAM			
ESVD			
EKJS			
MKJS			

Table 2: Hearing types that a judge can perform in the week before a holiday.

Table 3 shows the hearing types that have limitations in the week after the holiday. It should be noted that this applies from the third working day after the holiday. In reality, this rule – to not have a hearing on the first two working days – is not followed up because of the need to plan hearings in time. Also here, the coloured cells indicate impossible combinations.

Hearing type	**Length of holiday (weeks)**		
	1	**2**	**>2**
EKJS			
MKJS			

Table 3: Hearing types that a judge can perform in the week after a holiday.

2.1.2.3 Capacity

Judges and clerks are trained per case type. This means that not all judges and clerks are interchangeable for all case types. In general, there are two groups: the Family judges and clerks; and the Youth judges and clerks. On the contrary, case types such as G&O and VZ are performed by both family and youth judges and clerks. *Figure 8* shows the maximum number of cases that can be worked on per week per case type. It should be noted that this does not mean that this number of cases per case type is achievable every week, but it is the maximum weekly capacity.

There is little similarity between the maximum case capacity of judges and clerks. For most case types there is judge undercapacity compared to the clerks, showing once again that the resource scarcity is located at the judges.

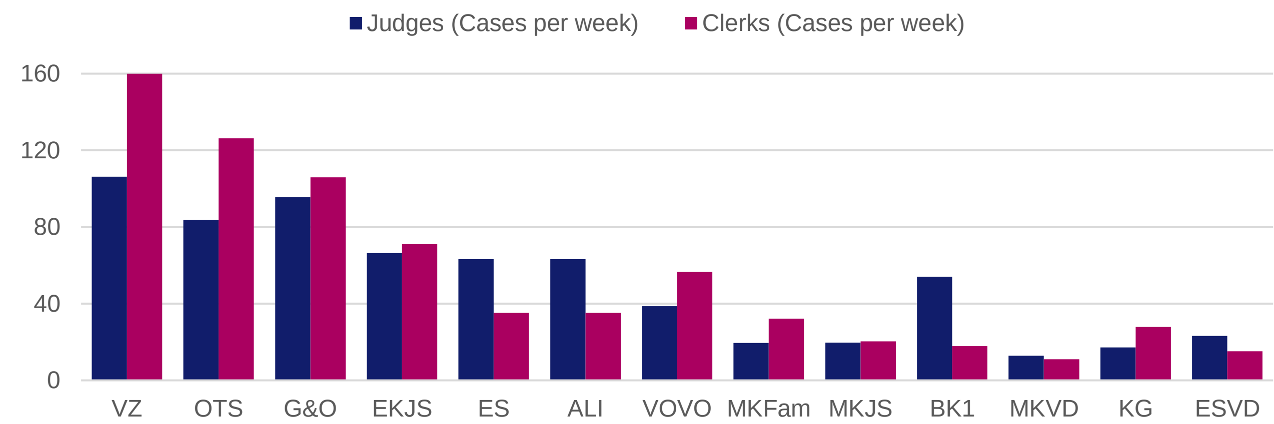


Figure 8: Maximum weekly case capacity of F&Y judges and clerks per case type (1-1-2025).

2.1.3 Case types per hearing type

Several case types are grouped in the same hearing type because of similarities in case type contents and overlapping specialisations of judges. *Figure 9* shows this grouping of cases. The eventual distribution of the different case types within a hearing type is determined by the case planner.

Figure 9: Case types per hearing type for Family (top row) and Youth (bottom row).

2.2 The share of buffer and idle time

Data from the scheduling software of F&Y is used to answer the question *What is the current share of the buffer and idle time in the judges' schedules?* Furthermore, we determine F&Y's resource scarcity by comparing the utilisation of the judges, clerks, and court rooms. The utilisation is determined with the following formula:

$$Utilisation = \frac{\#\ hours\ spent\ on\ primary\ activities}{\#\ hours\ allocated\ for\ primary\ activities}$$

The primary activity hours, as explained in *Section 1.2.2.2*, are the hours that should be spend on primary activities such as cases and hearings.

The data used is derived from the 2024 schedule (1/1/2024 – 31/12/2024). One year is used because of staff alterations. It is not possible to calculate the *primary hours utilisation* of staff that left or joined F&Y in the middle of the year over the time that they worked at F&Y, since the exact dates of joining or leaving are not known. Furthermore, the 2024 schedule also provides data about the most recent, thus relevant, situation of F&Y.

2.2.1 Judge utilisation

Figure 10 shows two boxplots with the judges’ utilisation. The left plot ($n = 20$) is concerned with the utilisation of all judges working at F&Y location Haarlem. The right plot ($n = 15$) contains only the judges that worked at F&Y Haarlem for the entirety of 2024 (i.e. judges that did not join or leave section F&Y throughout the year). It also excludes judges that were not able to work for a longer period of time because of varying circumstances (e.g. long-term illness, burnout, sabbatical). The reason to not include all F&Y judges is to give a clear picture of the realised utilisation caused solely by the schedule, and not by unforeseen circumstances which cannot be (directly) attributed to the schedule.

The average *primary hours utilisation* of all F&Y Haarlem’s judges is 64.1%. The *primary hours utilisation* of the judges that have been able to work for most of the year is 74.0% (with a minimum of 25% and a maximum of 91%). The deviating utilisation per judge can be explained by the arrangement of primary and secondary activities. The planning of certain secondary activities can lead to sizable planning gaps (idle time) between hearings. Furthermore, it can be that the primary hours factor of a judge is outdated; thus, not representing the current distribution between primary and secondary hours, thereby overestimating or underestimating the *primary hours utilisation*.

Because the schedule provides a *primary hours utilisation* of 74.0% on primary hours, we assume that 26.0% of the time is unscheduled in the scheduling software, meaning that more than a quarter of judges’ primary hours is scheduled as buffer and idle time.

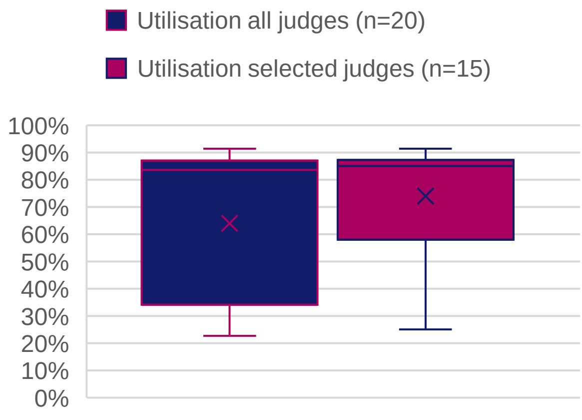


Figure 10: Box plots of primary hours utilisation of F&Y Haarlem’s judges (2024).

2.2.2 Clerk utilisation

The *primary hours utilisation* of F&Y Haarlem’s clerks is plotted out in two box plots (*Figure 11*). The same exclusion criteria are applied as to the judges. For all Haarlem clerks, the utilisation of the primary hours is 68.0%, and for the clerks that pass the exclusion criteria 69.9%.

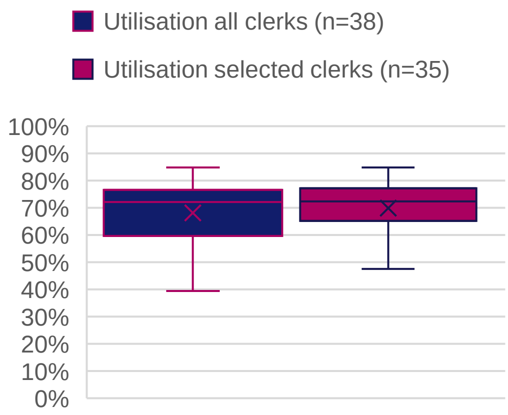


Figure 11: Box plots of primary hours utilisation of F&Y Haarlem's clerks (2024).

2.2.3 Courtroom utilisation

Figure 12 shows the courtroom utilisation of Court of Law Haarlem divided per EK and MK set up and the possibility to be secured. In 2024, 4,953 hearings took place at the Court of Law Haarlem. The utilisation of the courtrooms is calculated similarly to the judges and clerks utilisation. The allocated primary hours of courtrooms are 40 hours per week (five days of eight hours).

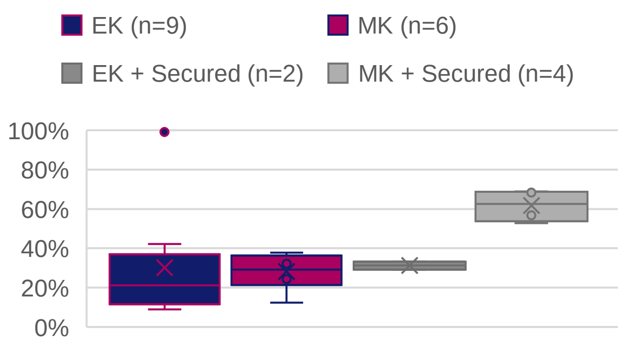


Figure 12: Utilisation of Haarlem's courtrooms divided in EK, MK, and optional security.

2.2.4 Bottleneck

For F&Y Haarlem, 5,206 primary hours of judges are unscheduled annually, 18,812 of clerks, and 28,310 of courtrooms (*Figure 13*). The discrepancy between time allocation distribution of hearing parts (i.e. preparation, hearing, finalisation) between judges and clerks is shown by the predicted needed times in *Appendix A.2 Time prediction*.

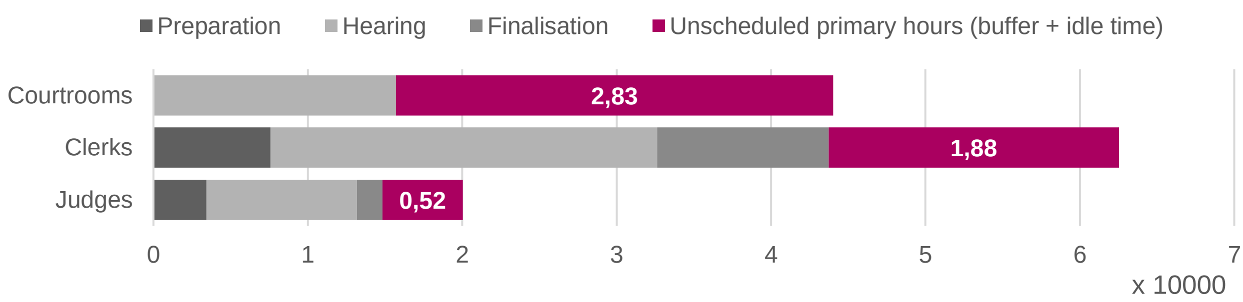


Figure 13: Total primary hours distribution of F&Y location Haarlem in 2024.

To predict the size of the idle time, we assume that the required buffer is not an entire daypart (4 hours) or larger, since that would mean that the buffer is the same size as, or larger than, the *predicted needed time* to finalise a hearing. If this is true, it means that the *predicted needed time* to finalise a hearing is half of what is needed in reality – which is assumed to be too deviant. With this assumption in mind, we calculate the range of the total buffer size of the judges:

$$0\ hours < buffer < finalisation\ time$$

$$0\ hours < buffer < 1{,}654\ hours$$

With this assumption in mind – that the buffer cannot be equal to or larger than the finalisation time – it means that the size of the idle time must fall within the following range, where *unscheduled primary hours* is abbreviated in uph:

$$\big(uph - MAX(buffer)\big) < idle\ time < \big(uph - MIN(buffer)\big)$$

$$3{,}552\ hours < idle\ time < 5{,}206\ hours$$

This means that between 18% and 26% of judges’ primary hours is idle time.

Applying the same assumption to the clerks yields that between 7,715 and 18,812 hours of the primary hours is idle time. This relates to between 12% and 30% of the primary hours.

Idle time can come to existence because of the time between the buffer and a hearing block break, as *Section 1.2.2.4* showed. It can also originate from secondary activities (i.e. meetings and courses) that are scheduled inconveniently on a day, creating idle time; or holidays which create additional idle time before or after them (*Section 2.1.2.2*).

Figure 14 shows the percentual distribution of primary hours for judges, clerks, and courtrooms.

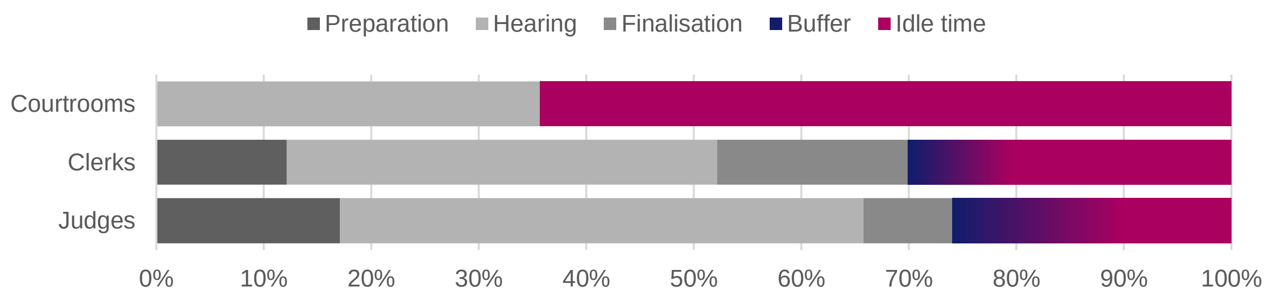


Figure 14: Distribution in percentages of total primary hours of F&Y location Haarlem in 2024.

With this information, together with that the maximum utilisation of the clerks is 85% (91% for judges), we assume that the resource scarcity is not located at the clerks of F&Y Haarlem. Furthermore, we assume the resource scarcity to be located at the judges, since clerks require less experience, they are available from a larger pool to hire. Moreover, the courtrooms are available in abundance.

Since the hearing schedule is made starting with the judges (*Appendix B.1 BPMN*), it supports the assumption that most attention of the hearing scheduling team is paid to the optimal scheduling of judges. Clerks are scheduled when a judge has been found, meaning that the clerks’ schedule is less prioritised compared to the judges’ schedule. The effect on the clerks’ *primary hours utilisation* is not significantly worse due to the scheduling after the judges. Therefore, we assume that reducing idle time of the judges will not have further deteriorating effects on the idle time of the clerks.

2.3 Case input, output, and backlog

Here, an answer is sought for the earlier proposed question: *What is the case input, output, and backlog of F&Y?* The case input is the demand that exists for F&Y. These are the cases that enter the system. In this section, a detailed description is given about the evolution and current state of the case input of F&Y. The case output is comprised of all cases that have been finished. The case backlog consists of all cases that are in the system.

2.3.1 Case input

Figure 15 shows the evolution of the total weekly case input of F&Y Haarlem. Available data starts in 2019. Halfway through 2020, the total case input decreases; which can be explained by changes in law which reduced the need for certain case types (de Rechtspraak, 2024a). An outlier is around new year between 2019 and 2020. At the end of 2019, case demand is outlying high, and at the beginning of 2020 it is outlying low. The total case input shows a light decrease year-on-year; stabilising from 2022 onwards. Additionally, a moving average of 26 weeks (half year) is added to show this stabilisation.

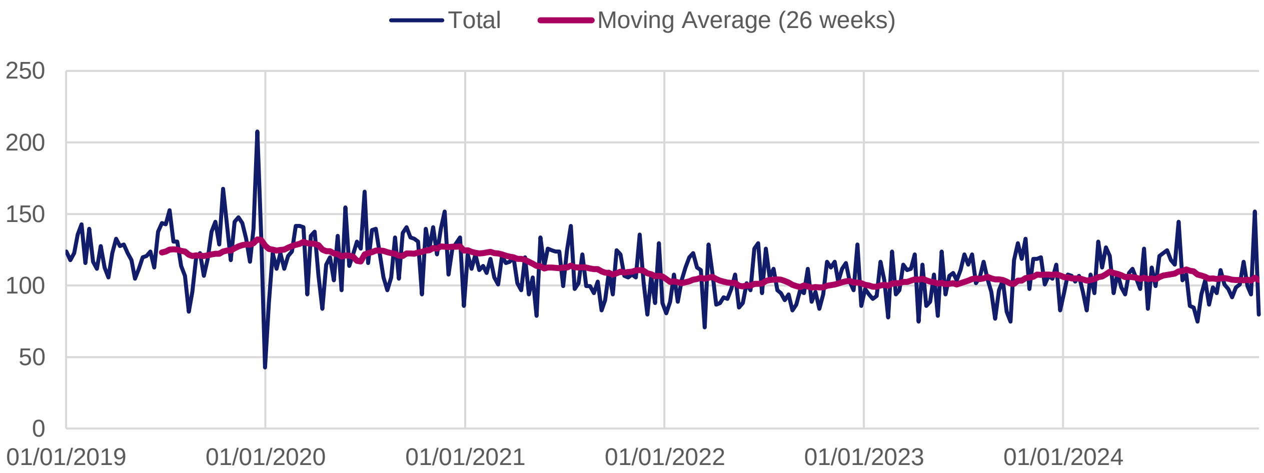


Figure 15: F&Y Haarlem's total weekly case input (n=34,898; 2019-2024).

Appendix B.2 Annual case input shows the annual case input level per case type. The case types show relatively stable arrival rates.

Figure 16 shows boxplots of F&Y Haarlem's total weekly case input and additionally also divided in a Family and Youth side. Weekly case input per side is relatively equal.

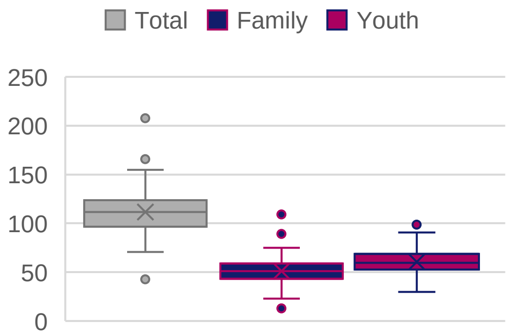


Figure 16: F&Y Haarlem's weekly total case input in boxplots (n=34,898; 2019-2024).

Appendix B.3 Weekly case input per case type shows boxplots that show the weekly case input per case type. Also, the standard deviations of the weekly case input over the years are present, showing little fluctuations in case input.

Figure 17 depicts the percentual year-on-year mutation for the Family and Youth side, as well as for the total case input. From 2019 onwards, case input decreased. From 2021 onwards, case input shows a stagnation in decrease, and from 2023 the case input increases year-on-year.

We assume that the demand does not show significant trends in change over the years, except for a small general decrease. Therefore, we assume that demand – case input – is constant. Even though there are some fluctuations in demand change year on year, these are limited and do not show a cohesive pattern, thus assumed negligible. Furthermore, the sample size – the years – is not extensive enough to determine a pattern with certainty.

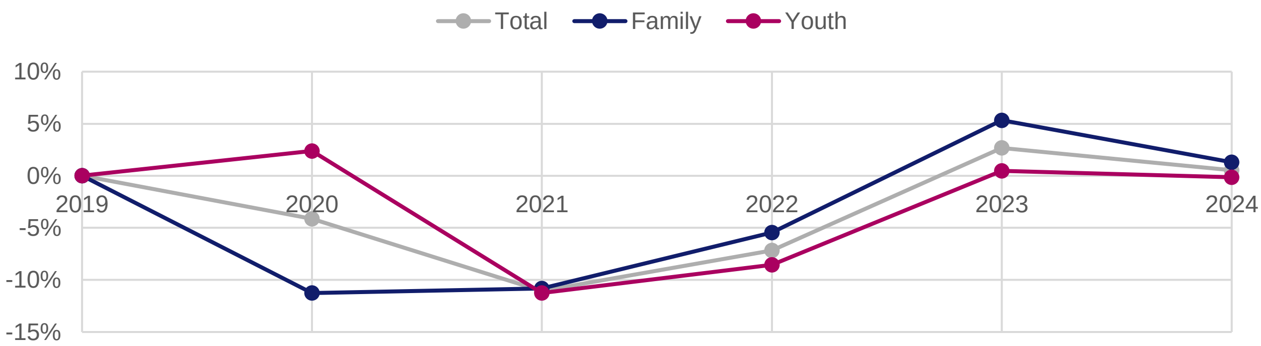


Figure 17: F&Y Haarlem's year-on-year total, Family, and Youth case input mutation (n=34,898; 2019-2024).

The case mix of the case input of 2019 up and until 2024 of F&Y Haarlem is shown by *Figure 18*. The Family case types are highlighted in purple; the Youth case types in pink. They are sorted on occurrence. We assume consistency of case mix over the years because of the relatively stable case input per case type as *Appendix B.3 Weekly case input per case type* shows visually and by the small standard deviations of case input.

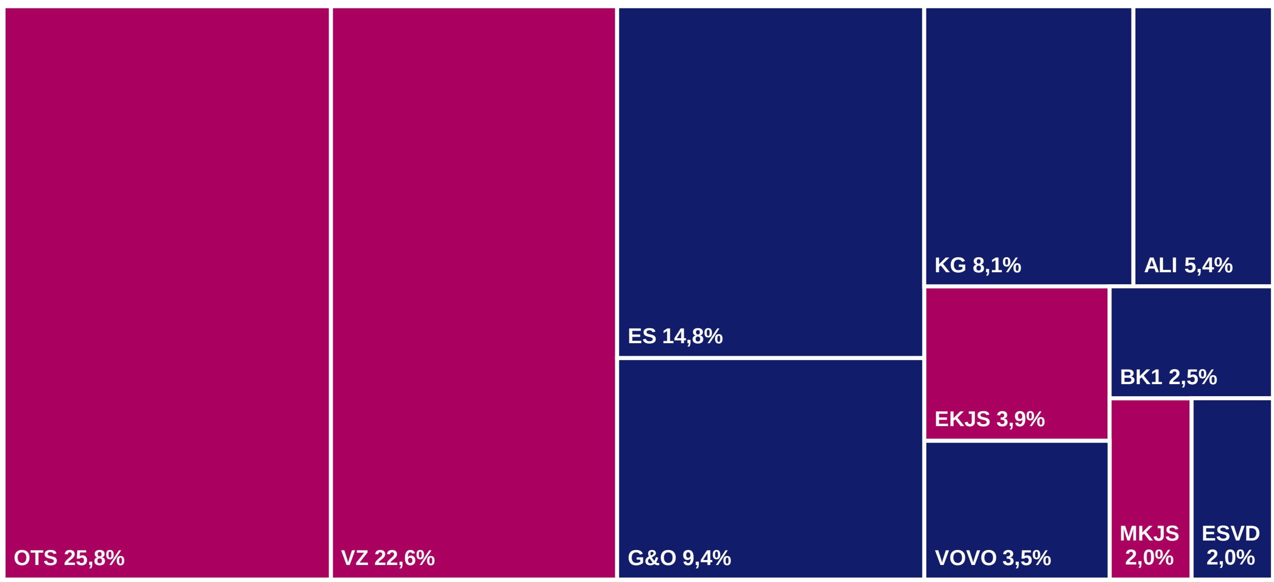


Figure 18: F&Y Haarlem's case mix (n=34,898; 2019-2024).

2.3.2 Case output

The exact numbers of case output per week nor per case type are available. Therefore, to determine the service rate, we make an approximation based on the *theoretical judge capacity*. We do this in *Chapter 4*. In reality, the case output will be lower than the case input, because of the existence of case backlog.

2.3.3 Case backlog

As a consequence of the lack of digitalisation of certain cases – which solely exist on paper – and inaccurate administration, the exact size of case backlog is unknown. Cases are scattered around drawers, offices, and courts; which makes determining the exact backlog size nigh impossible for Court of Law Noord-Holland.

Predictions have been made to determine the exact size of case backlog – these are shown by *Figure 19*. F&Y's hearing scheduling team is counting cases that still need to come on a hearing, setting the lower bound of the case backlog. According to the case administration system, the backlog numbers are significantly greater. However, due to inaccurate administration not all of these cases need to come on a hearing; thereby, providing an upper bound to the case backlog. Therefore, the real case backlog must be between the already identified (by the hearing scheduling team) and the maximum case backlog (case administration system).

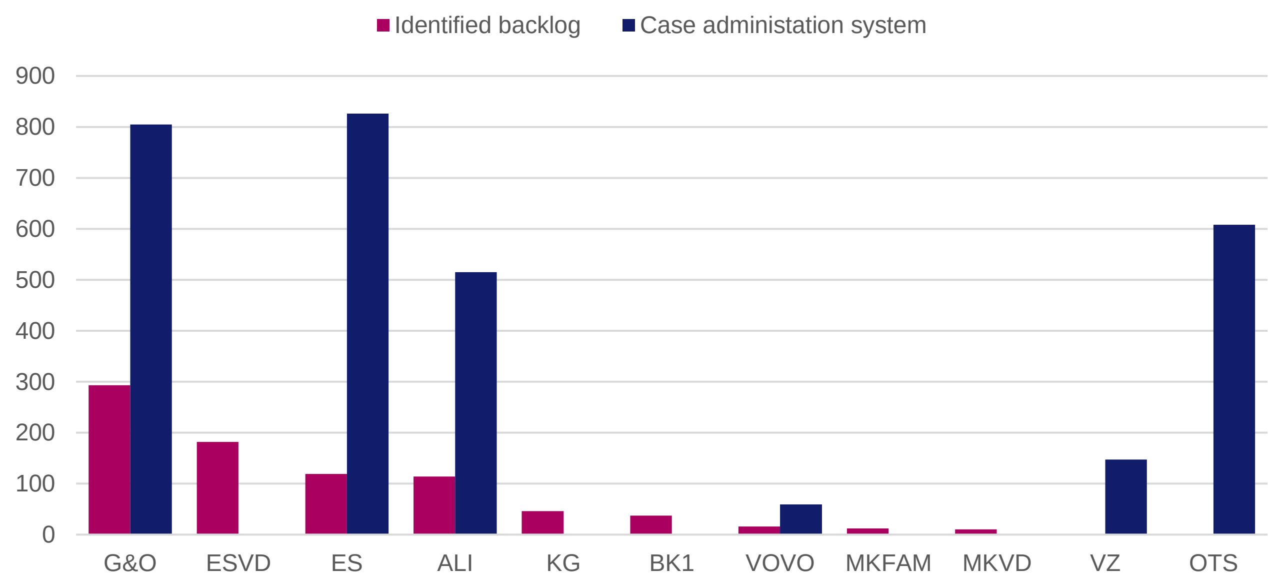


Figure 19: F&Y's case backlog per case type (1-1-2025).

Figure 20 shows the case backlog according to the case administration system and identified by the scheduling team. It can be seen that there is no existent case backlog at the Youth side according to the identified backlog.

An explanation for the nonexistence of case backlog at the Youth side is that these case types have stricter deadlines that need to be enforced (i.e. expiration of OTS verdicts); resulting in a transfer of capacity from the Family side to these case types which in turn causes undercapacity at the Family side (de Rechtspraak, 2021b).

The total existent case backlog is between 829 and 2,960 cases as of 1 January 2025.

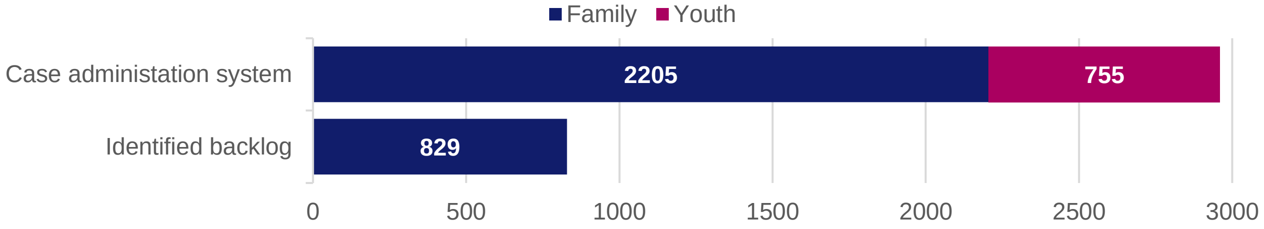


Figure 20: F&Y's case backlog per Family and Youth side (1-1-2025).

2.4 Conclusion

In this chapter, its respective question – *How does the hearing scheduling process work currently?* – is answered in the following ways:

The process of the creation of the hearing schedule is determined by means of a BPMN. Also, the difference between case types and hearing types is explained, as well as the constraints related to scheduling at F&Y.

An approximation of the size of the buffer and idle time is made. The idle time for judges ranges between 18% and 26% of the primary hours. Moreover, we showed that F&Y Haarlem’s resource scarcity is located at the judges – not at the clerks or courtrooms.

Case input of F&Y Haarlem is explored in general, per Family side, per Youth side, and per case type. No case output data is available. The backlog numbers are not known exactly, but a range for the case backlog is determined. The case backlog is located heavily at the Family side.

3 Literature review

Chapter 3 broadens the knowledge of the needed techniques and the awareness of earlier conducted research as elaborated on in *Section 1.3.1.4*. *Section 3.1* develops understanding of linear programming techniques which are used for scheduling. *Section 3.2* is concerned with the discovery of queuing techniques in the scheduling process. For both sections; first, literature is discovered which is applied in the judicial sector; then, literature which is applied in the service sector, but which can also be applied to F&Y. The linear programming section (*3.1*) also looks at the makespan phenomenon used in production management.

3.1 Linear programming techniques

3.1.1 Judicial system

Linear programming – including its subcategories (i.e. (mixed) integer linear programming) – is a limited researched topic within the judicial system. Nevertheless, there is some literature around on this topic concerning the judicial system which focusses on judge allocation (Gupta & Bolia, 2024), court allocation (Teixeira et al., 2019), and hearing scheduling (Brooks, 2012; Jennings, 1971).

Brooks (2012) formulates an integer programming model that is used to make a hearing schedule, which also assigns judges to the scheduled hearings. It must be noted that Brooks' model is aimed at reducing scheduling costs instead of improving the hearing schedule or judges' utilisation.

Albeit some time ago, Jennings (1971) developed a linear programming model for the *New York City RAND institute* that encapsulates costs, fairness, legal constrains, and backlog, among others. Despite the extensiveness of the model, it is solely concerned with the hearing scheduling.

3.1.2 Service sector

In an analysis of different approaches to the *University Course Timetabling Problem* (UCTTP), Babaei et al. (2015) find that applying (integer) linear programming to the UCTTP – outside faculty specific application – will hardly yield optimal solutions for education institution broad application due to the complexity of the problem (Babaei et al., 2015).

For the *Athens University of Economics and Business*, Dimopoulou and Miliotis (2001) group courses together in blocks of four hours that share similarities. By aggregating the courses, they reduce the number of variables. In that way they are able to find an optimal solution using linear programming (Dimopoulou & Miliotis, 2001).

When a maximization of capacity utilisation is sought, it remains important to have slack in a schedule when durations of activities are unknown to "make a surgery schedule more robust against overtime" (Hans et al., 2008). By means of utilizing the *portfolio effect* (Markowitz, 1991), they can minimize the slack while keeping the schedule robust.

In healthcare scheduling – nonetheless also applicable to the judicial system – it is important to keep Su & Zenios' *equity-efficiency* dilemma in mind. The prioritization policy of patients influences the efficiency and the equity distributed among them (Su & Zenios, 2004).

The found literature is concerned with the scheduling of activities of predetermined lengths on the operational and tactical level. However, we aim at a change on a high tactical level, which is underdeveloped by existing literature, showing the existing gap in the literature.

3.1.3 Makespan

The makespan is defined as the time between the begin and end of a production process. The problem at F&Y shows parallels with batching problems in production management. Relevant batching problems are solved by minimising the makespan by determining the optimal batch size (İnce et al., 2024; Ji et al., 2023).

Literature concerning batching is centred around *s-batching* and *p-batching*. They represent the either serial or parallel processing of jobs on a machine, respectively. In *s-batching*, one job can be processed at a time; whereas with *p-batching* several jobs are processed at the same time by a machine (*Figure 21*) (Fowler & Mönch, 2022; Potts & Kovalyov, 2000).

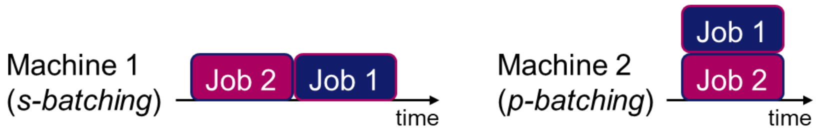


Figure 21: S-batching and p-batching visualisation.

In F&Y's case – even though the hearing schedule itself represents *p-batching* – from a judge's perspective, *s-batching* represents the situation at hand best since cases are treated after each other in a hearing. A judge is not able to work on multiple hearings at the same time; hence, *p-batching* does not fit the judge's perspective.

Within *s-batching*, the batch processing time is calculated as the sum of the processing times of all jobs within the batch (Huang, 2025; Paulus et al., 2009).

In a literature review of flow-shop scheduling, İnce et al. (2024) provide an extensive arrangement of flow-shop scheduling models with differing characteristics based on the *three-field notation* (İnce et al., 2024).

Graham et al.'s *three-field notation* is a notation form for scheduling problems. It attains the following structure: $\alpha \mid \beta \mid \gamma$, where α denotes the machine environment, β the characteristics and constraints of a job, and γ the objective function (Graham et al., 1979).

F&Y has parallel machines (judges), which are interchangeable – assuming that both sections (Family and Youth) are mutually exclusive and thus have different processes. There are precedence constraints: A hearing can only start when the preparation has finished, and finalisation when the hearing has finished. Then, there are batches of cases in a hearing which can differ per case type. The goal is to minimize the makespan – completion time – to minimize the idle time in the hearing schedule. Following the *three-field notation*, F&Y's job scheduling attains the following notation (Graham et al., 1979; Lawler et al., 1993):

$$P_m \mid prec, batch(j) \mid C_{max}$$

Table 4 compiles found literature showing parallels with F&Y's problem according to the *three-field notation*. The boldfaced characteristics indicate similarities with F&Y.

Authors	α	β	γ
(Boccia et al., 2013)*	$\boldsymbol{P_2}$	$\boldsymbol{batch}$	$\boldsymbol{C_{max}}$
(Huang, 2025)	$\boldsymbol{P_m}$	$m - batch, r_j, w_j, s_j, c$	$\boldsymbol{C_{max}}$
(Hwang et al., 2012)*	$F2$	$\boldsymbol{prec}, idle$	$\boldsymbol{C_{max}}$
(Jinsong et al., 2009)*	$\boldsymbol{P_m}$	$\boldsymbol{prec}$	$\boldsymbol{C_{max}}$
(Kozik & Rudek, 2018)	1	$r_j, s_{ij}, \boldsymbol{prec}, ma, p_j(v_j)$	$\boldsymbol{C_{max}}$
(Li et al., 2022)*	1	r_j	$\boldsymbol{C_{max}}$
(Pandey et al., 2023)*	R	$pmtn, \boldsymbol{batch}$	$\sum T_j$
(Rocholl & Mönch, 2024)	$FF2$	$F_s, \boldsymbol{s - batch}, r_j, s_j, q_s$	TWT
(Shen & Buscher, 2012)*	J	$\boldsymbol{batch}, s_{ij}, r_j$	$\boldsymbol{C_{max}}$
(Xie et al., 2025)	1	s_j, B, p_j	$\boldsymbol{C_{max}}$
(Yang et al., 2022)	1	$p - batch, v_i, incompatible$	$\sum w_i C_i$
(Zou et al., 2012)*	$\boldsymbol{P_m}$	$\boldsymbol{prec}, chain$	$\boldsymbol{C_{max}}$
**Three-field notation* not explicitly mentioned; assumptions made based on descriptions in the respective article (Graham et al., 1979; Lawler et al., 1993).			

Table 4: Three-field notations of found literature.

In an $FF2 \mid F_s, s - batch, r_j, s_j, q_s \mid TWT$ model, Rocholl and Mönch (2024) aim to minimize tardiness – the lateness of job completion, in additive manufacturing since tardiness is considered to be a performance measure. This is different from F&Y since at F&Y the aim is not to minimize tardiness, but to minimize the makespan. Nonetheless, they propose a MILP model to solve the flow shop problem. Their research discusses a problem of several stages that jobs have to pass through. Only jobs of the same family can be batched together (Rocholl & Mönch, 2024). This research shares characteristics with F&Y, since cases have to pass through several stages (i.e. preparation, hearing, and finalisation), and in a hearing only similar case types (jobs) can be batched together. A disparity with F&Y is that Rocholl and Mönch have set up times, a phenomenon that is not applicable to F&Y.

Although the *three-field notation* is commonly used within scheduling problems, problems following exactly the same characteristics as F&Y's are not found.

3.2 Queuing techniques

In this section, an answer is sought for the two questions related to the search of queuing theory techniques as defined in *Section 1.3.1.4*.

Kendall's notation – named after its creator – describes the main characteristics of queuing systems. The general notation attains the following parameters: $A/S/c$; where A is the arrival rate, S the distribution of the service time, and c the number of servers (Kendall, 1953). Queues with a Poisson distributed interarrival time are depicted with an M. If the service time is distributed exponentially, an M is written. If it is generally distributed a G is written. For example, a queue with a Poisson distributed arrival rate and generally distributed service time with one server is depicted as $M/G/1$.

3.2.1 Judicial system

Concerning the *Supreme Court of India*, Bakshi et al. (2025) show how queuing techniques can be applied within the judicial system to reduce delays in delivering justice. The two main factors which they address are: the delay; and the cases awaiting adjudication (i.e. backlog). According to the authors, the factors are connected by *Little's law* (Stidham, 1974):

$$Expected\ backlog = expected\ arrival\ rate * expected\ delay$$

Using a *case-management queue*, they show that they can reduce delay by making its drivers (i.e. number of postadmission cases in the queue) insightful. The *case-management queue* is distinctive in that it has multiple *service episodes* (i.e. hearings), but with the *same server* (i.e. judge panel) (Bakshi et al., 2025). Although they are not able to precisely determine the utilisation and other characteristics of the queue – due to the complexity of *case-management queues* – they can derive outcomes based on simulations. Here, they find that delays can be reduced by up to 65% when the drivers are known and a queuing policy is tailored to these.

Bakshi et al. refer to research conducted by Bray et al. (2016) in the *Roman Labour Court of Appeals*. The system is concerned with case scheduling – instead of hearing scheduling – by means of a *multiarmed bandit* problem. Here, the judge's decisions are determined by the objective, which is to maximize the number of cases finished. They bring about a solution which has profound impact (i.e. case flow time reduction of 111 days). Bray et al. argue that a relatively simple solutions has great effect in this environment where there is no operations research around. The authors provide a fitting description of the Italian judiciary: "This environment [the Italian judiciary] is ripe for operations research – it is critical, complex, and wasteful" (Bray et al., 2016).

During the application of their queuing policy in the *Roman Labour Court of Appeals*, Bray et al. noted that a queuing policy should "(i) be clearly beneficial, (ii) preserve judicial autonomy, (iii) not increase workloads [of judges], and (iv) be easy to understand and implement [by schedulers]" to make it a success. To convince the relevant stakeholders (e.g. judges, clerks, schedulers, management) that the policy improves the situation at hand, the queuing policy should include these characteristics (Bray et al., 2016). Since this holds for the case schedule, the assumption is made that this can be extrapolated to the hearing schedule as well.

3.2.2 Service sector

In the service sector, more research has been conducted on applying queuing theory in scheduling. Banks (Cho et al., 2017), hospital surgeries (Ferdinandes et al., 2017), outpatient systems (Sheikh et al., 2013), and organ allocations (Su & Zenios, 2004) are described as $M/M/s$ queues.

On the other hand, some outpatient services (Brahimi & Worthington, 1991) and prisons (Sonenberg et al., 2024) are described as $M_t/G/S$ queues.

In the prison system of England and Wales, Sonenberg et al. (2024a) try to predict queue lengths based on the prison's demand. For the fact that balking is not possible in the prison system (similar to the judicial system), the authors assume that *infinite server queuing models* are applicable. The prison system is described as a "$M_t/G/\infty$ queue with a non-homogeneous Poisson arrival process." The model of Sonenberg et al. predicts short and long-term behaviour of the system. Like Bray et al. (2016), the authors are convinced that a relatively simple application of queuing theory can assist in the "approach to potential policy changes in the real world", albeit that the model is made with a lot of assumptions due to missing and incomplete data (cf. F&Y) (Sonenberg et al., 2024).

Abundant literature concerning queuing theory in scheduling in the service sector is around. Most queues are identified as either $M/M/s$ or $M_t/G/s$ queues. In the judicial system however, limited research is conducted on the topic of queuing theory. For the judicial system, it is concluded that even relatively simple implementations of queuing theory can yield major benefits and improvements in terms of delays, waiting time, and backlog reduction. Only, the implementation of policies created provides possible challenges.

3.3 Conclusion

In this chapter, we sought an answer to the question *What literature exists for scheduling?* Knowledge is broadened about linear programming and queuing techniques in both the judicial system and the service sector.

Applying linear programming in the judicial system is still in its "infancy." In the service sector, a widespread problem is the number of variables that prohibits the calculation of the optimal solution. By means of aggregating the number of variables, the calculation time is reduced, and it is made possible to find an optimal solution. One should be aware of the *equity-efficiency* dilemma. The literature concerning linear programming at a high tactical level is underdeveloped.

The *three-field notation* of F&Y is determined ($P_m \mid prec, batch(j) \mid C_{max}$) and related systems are examined. No system is found that is exactly similar to F&Y.

By making drivers of delay insightful and tailoring a relatively simple queuing theory to it, major improvements can be made. Provided that the implementation of the queuing policy follows certain rules. It is difficult to use queuing theory to precisely predict outcomes because many assumptions have to be made to fill up uncertainties.

[This page intentionally left blank]

4 Solution development

Chapter 4 develops the solution for F&Y. *Section 4.1* answers the earlier proposed question – *How does the hearing length influence the idle time length between hearings?* – by means of a mathematical model to determine the influence of the hearing length on the idle time. *Section 4.2* identifies the queue parameters and characteristics of F&Y to answer the question *What are the parameters of F&Y's queuing system?*

4.1 Mathematical model

To provide an answer to the question *How does the hearing length influence the idle time length between hearings?* – we propose a Mixed-Integer Linear Program (MILP), since both the input and output are based on integer input. This model is used to be able to determine the influence of the hearing length on the idle time. The full model is located in *Appendix C.1 Mathematical model*. Here, the model's key elements are explained.

Figure 22 visualises the scheduling of one hearing process (j). A hearing process is comprised of all hearing parts (i.e. preparation, hearing, and finalisation).

The decisions taken are the starting moment of the preparation (P_{tij}) of the hearing process, the starting moment of the hearing (H_{tij}) of the hearing process, and the number of cases to schedule together (x_i) in the hearing process of hearing type i.

The size of x_i determines the time needed for the preparation (p_i), hearing (h_i), finalisation (f_i), and buffer (β_i) of all cases in hearing process j.

The total number of hearing processes (j) to schedule is determined both by the case input (ε_i) and the hearing length (x_i).

Q is the makespan of all hearing processes (j) and aimed to be minimised.

The extensive explanation of the decision variables, together with their constraints, is described below.

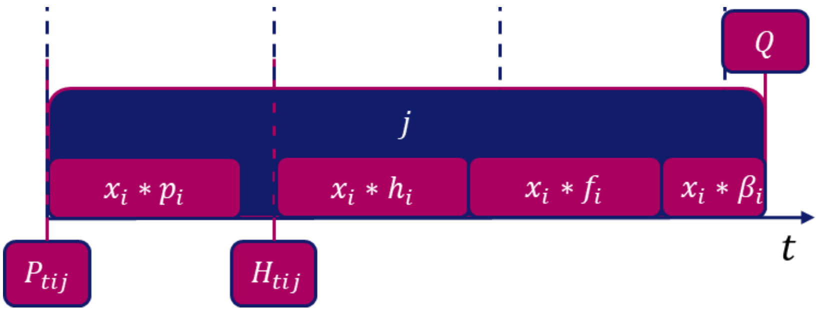


Figure 22: Visual representation of scheduling hearing process j.

The model represents a single judge – working fulltime – with a primary hours coefficient of one; meaning that all work hours in the week (40) are spent on cases and hearings. The model also assumes that the judge is able to perform all case types of either the Family or Youth side without preferences.

The main decision to take is the number of cases to schedule together in a hearing process per hearing type. In the model, this decision is denoted with the integer variable x_i where the index i denotes the hearing type. This decision variable attains a value from the following set: $x_i \in \{1,2,3,4\}$. 1 case per hearing is the lower bound since it is the minimum number of cases to schedule on a hearing; 4 is the upper bound since more than 4 cases on a hearing has negative implications on the quality of justice delivered according to research conducted and prescriptions by de Rechtspraak (de Rechtspraak, 2021b; Staats & Gino, 2012; Tan & Netessine, 2014). Also, when more than 4 cases are scheduled on a hearing, the hearing block break is overwritten for most case types (*Section 1.2.2.4*). In production management terms, the hearing length decision (x_i) is comparable to *s-batching* as *Section 3.1.3* describes.

The cases that need to be scheduled (ε_i) need to be scheduled in hearing processes (j). The number of hearing processes j needed is dependent on ε_i and the number of cases to schedule together in a hearing (x_i). The more cases are scheduled together in a hearing, the fewer hearing processes are needed.

To determine the start and finishing times of these needed hearing processes (j), the binary decision variables P_{tij} and H_{tij} are used to determine whether the preparation and hearing start at time t of hearing type i of hearing process j respectively. The finalisation of a hearing process is not scheduled separately since we assume that it can begin directly after completion of the hearing. Following *Section 3.1.2*, this saves computational time by reducing the number of variables without affecting the outcomes.

The model minimises the makespan of all cases that need to be scheduled by taking the optimal decision of x_i. When the makespan is minimised, the finishing time of the last hearing process (j) needed is reduced as much as possible. By minimising this time, the idle time between hearings is also minimised. This objective function is denoted as $Min\ Z = Q$, where Q is subject to a constraint which facilitates the minimisation of the makespan – and consequently idle time. Q attains the earliest finishing value of the last hearing process $(t + (\,h_i + f_i + \beta_i) * \psi_{tij})$. Inherently, the idle time is minimal when the makespan is also minimal. The auxiliary variable ψ_{tij} is used to determine whether the number of cases x_i is scheduled or not:

$$\sum_{t=1}^{T}(t + (\,h_i + f_i + \beta_i) * \psi_{tij}) \leq Q \quad \forall i \in I, \forall j \in J$$

The hearing parts (preparation and hearing) need to follow a specific order. For instance, the hearing cannot take place if the preparation of that hearing process has not been completed. Precedence constraints are used to force this order. The following constraint forces the time of starting a hearing (H_{tij}) of hearing process j to be greater than the time of completing the preparation (P_{tij}) of hearing process j. ω_{tij} is used to determine whether x_i cases are scheduled on P_{tij}:

$$\sum_{t'=1}^{T}\left(t' * H_{t'ij}\right) \geq t * P_{tij} + p_i * \omega_{tij} \quad \forall t \in T, \forall i \in I, \forall j \in J$$

Also, the preparation of the next hearing process (P_{tij+1}) can only start at a time greater than the finalisation of hearing process j. ψ_{tij} is used to determine whether x_i cases are scheduled on H_{tij}:

$$\sum_{t'=1}^{T}\sum_{i'=1}^{I}\left(t' * P_{t'i'j+1}\right) \geq \left(t * H_{tij} + (h_i + f_i + \beta_i) * \psi_{tij}\right) \quad \forall t \in T, \forall i \in I, \forall j \in J\backslash\{|J|\}$$

To determine which hearing type (i) to assign to which hearing process (j), the binary decision variable δ_{ij} is introduced. This variable operates under the constraints that (a) there can be at most one hearing type assigned to a hearing process; (b) enough hearing processes should attain hearing type i to meet all case demand of i (ε_i); (c) and all hearing parts (preparation (P_{tij}) and hearing (H_{tij})) of hearing process j should be planned of the same hearing type i.

The main reason of the existence of idle time, is the hearing block break. Because of the rigid hearing block structure (*Section 1.2.2.4*), a hearing (H_{tij}) cannot cross the break between two hearing blocks (i.e. morning-afternoon break and day division). To enforce this hearing block break, a modulo operator is used. Since a modulo operator is not linear, it is rewritten to fit in a linear model (Pesant et al., 2024). The remainder of the modulo (r_t) of the starting hour (t) of a hearing is determined with $r_t = (t-1) - \theta * \gamma$ where r_t is subject to $0 \leq r_t < \theta$. t is the dividend and θ the divisor.

n_t represents the number of hours from t remaining until the next hearing block break. Based on this value, the choice is made whether a hearing can start at t or not. Since an if-statement is not linear, a *Big-M* constraint is used to decide whether a hearing can start based on the time left until a hearing block break (n_t) and the time needed to finish the hearing ($h_i * x_i$).

Figure 22 shows that the start of the hearing (H_{tij}) does not directly follow up on the end of the preparation. It would cross the hearing block break (dotted line) if it does, which is impossible. Since the hearing cannot directly begin after the finishing of the preparation causes idle time to exist.

$$n_t - h_i * x_i \geq -M^{\theta}\left(1 - \sum_{i'=1}^{I} H_{ti'j}\right) \quad \forall t \in T, \forall i \in I, \forall j \in J$$

4.2 Queuing system

This section answers the following question: *What are the parameters of F&Y's queuing system?* To answer this question, queuing theory is applied. Following *Kendall's notation*, as explained in *Section 3.2*, the important parameters are the arrival rate, the service rate, and the number of servers of the queue (Kendall, 1953). When these parameters are known, we can perform experiments and predict backlog and waiting time values under those experimental circumstances.

4.2.1 Arrival rate

Figure 23 shows boxplots of the arrival rates of cases grouped per Family and Youth side. On average, the arrival rate of Family cases is fifty-one per week, and sixty-one per week for Youth cases. The data is from 2019 up and until 2024. The outlying values on the Family side are already described by *Section 2.3.1*. The occurrence of these is exceptional and not representative of patterns in arrival rate deviation. Therefore, we assume that the average arrival rate represents the real arrival rate.

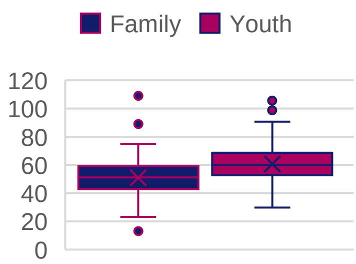


Figure 23: Arrival rate boxplots Family and Youth (2019-2024).

In accordance with *Section 2.3.1*, we assume that there are no significant trends in demand change. Based on the weekly case input, a histogram is created in *Figure 24*, showing the frequencies of the case input. Here, the previously discussed outliers are left out. Due to the shape of the right skewed histogram, we assume that a discrete *Negative Binomial* distribution depicts the arrival rate distribution of cases. An overlay of the expected frequency under the chosen *Negative Binomial* distribution is shown in the figure alongside the realised arrival frequencies.

In the literature (*Section 3.2.2*), the *Poisson* arrival rate is a common sight in the judicial system and the broader service system. However, for F&Y this is not viable since a *Poisson* distribution would not fit to the arrival rate because the mean and the variance are not equal (Meijer, 2023).

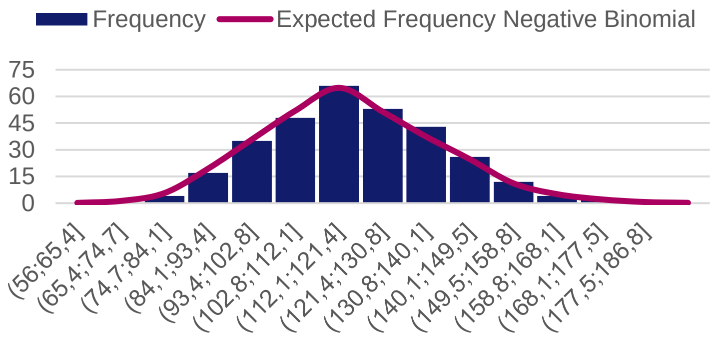


Figure 24: Weekly case input frequency (2019-2024) and Negative Binomial approximation.

In *Appendix C.3 P-P plot & PCC*, a P-P plot is located which shows that the *Negative Binomial distribution* is a suitable distribution for F&Y’s case input data. Furthermore, *Pearson’s correlation coefficient (PCC)* shows a strong correlation between the expected case input frequency under the *Negative Binomial* distribution and the observed case input values of F&Y.

A *Chi-square test* compares the observed frequencies with the expected frequencies of the *Negative Binomial* distribution mathematically, confirming whether the expected distribution fits the data. The *Chi-square test* can be executed when two datasets are independent of each other. A *Chi-square test* is conducted with the following formula; where O_i is the observed value and E_i the expected value according to the chosen distribution (Meijer, 2023):

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

If the value of the *Chi-square test* (χ^2) is smaller than the *critical value*, the difference between the observed and expected values is not statistically significant. The *critical value* is found to be 31.41 (from a *Chi-square* probability table), the value of the test 4.75. 4.75 is smaller than 31.41, hence the difference of the observed and the expected values is not statistically significant. Therefore, the *Negative Binomial* distribution is a plausible distribution for the arrival rate of cases.

Because not all cases need to come on a hearing, or need more than one hearing, the average weekly input of cases is adjusted by probabilities – determined by de Rechtspraak, based on historical data (*Figure 25*). One may argue that it is more accurate to determine the arrival rate distribution after applying these probabilities. However, whether a case needs to come on a hearing or not, it always enters the system. Therefore, the arrival rate distribution is determined based on the total case input.

For the EKJS and MKJS case types, the probabilities are unknown. Therefore, we assume that all incoming cases need one hearing. KG cases are short stream cases, thus by definition it is not possible to have multiple hearings. The average of the weekly input is deemed a valid value to use for these calculations due to the lack of outliers as shown in *Section 2.3.1*. Since not all cases come on a hearing, the arrival rate (*Figure 26*) is lower than the earlier identified case input.

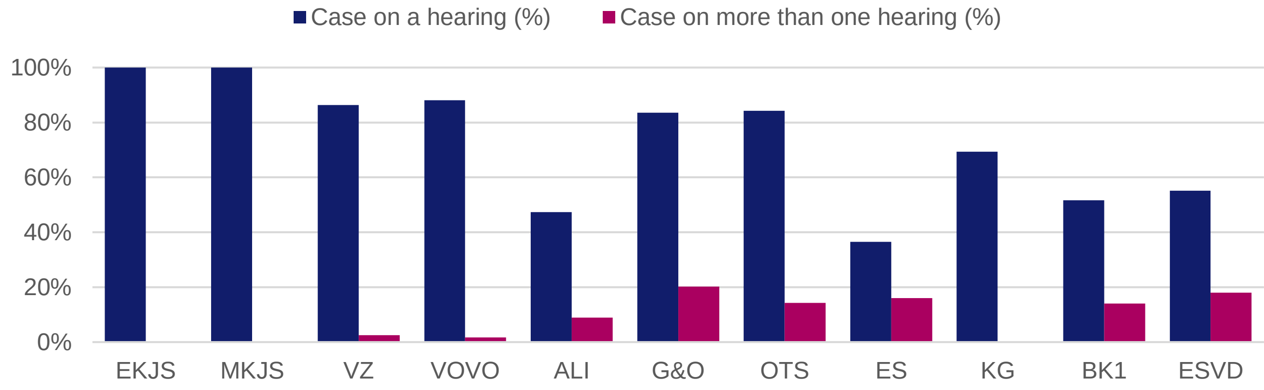


Figure 25: Probabilities of a case coming on a hearing and needing more than one hearing.

4.2.2 Service rate

The service rate (μ) is determined based on the theoretical capacity of the judges, and not on achieved case output for the reason of data unavailability of the latter. The *theoretical judge capacity* is deemed the best approximation of the service rate. We calculate the service rate with the following steps: First, the average case duration is calculated using the *predicted needed time* (*Appendix A.2 Time prediction*) and the case mix (*Section 2.3.1*). Then, after dividing the judges in a Family and a Youth side – because of judge specialisation similarities – the primary hours are added up per side and divided by the average case duration. This is the service rate (μ), which can be seen in *Figure 26* together with the average weekly input – arrival rate (λ) – of the Haarlem court (also grouped in Family and Youth sides).

The service rate is higher than the arrival rate; hence, the utilisation of primary hours is lower than one. An important fact that should be noted is that we assume mutual exclusivity between the Family and Youth side. Although that this is not necessarily the case in reality (i.e. G&O and VZ can be performed by some Family and Youth judges), we assume this since otherwise we cannot determine the distribution between the primary hours effectively.

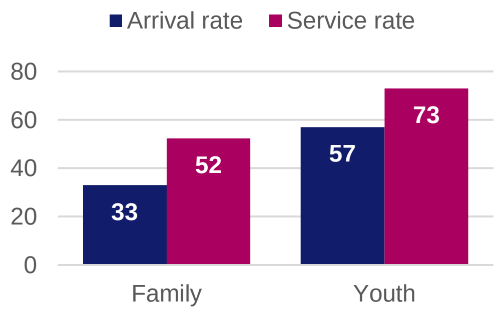


Figure 26: F&Y Haarlem's weekly arrival and service rate.

4.2.3 Servers

Since the judges are the bottleneck of F&Y, as determined by *Section 2.2.4*, they are used to determine the number of servers. As of 1 May 2025, sixteen judges work at F&Y Haarlem. eight of them are specialised in Family case types, eight in Youth case types.

Assuming that all Family judges are able to perform all Family case types; and that all Youth judges are able to perform all Youth case types, the utilisation rates of the primary hours of both the Family and Youth side are calculated (*Figure 27*). We do this with the following formula, where λ is the arrival rate, μ the service rate, and ρ the utilisation of the primary hours:

$$\rho = \frac{\lambda}{\mu}$$

The utilisation rates of the primary hours are less than 100%, meaning that all demand can be processed. However, due to the existence of case backlog, we argue that the service rate – based on the *theoretical judge capacity* – is not realised. In reality, the utilisation is higher than 100% because of backlog existence.

This implies that either the *theoretical judge capacity* is overestimated, or the needed time for cases and hearings is underestimated. Furthermore, the theoretical utilisation is higher than the realised utilisation of *Section 2.2.4*, reconfirming that the theoretical utilisation does not correspond with the realised utilisation.

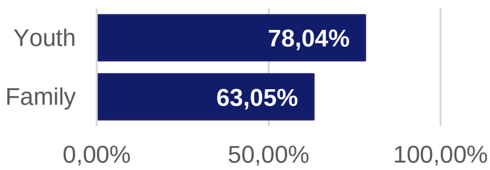


Figure 27: F&Y Haarlem's primary hours theoretical utilisation per judge.

Describing F&Y with *Kendall's notation* provides the following notation: $G/G/z$.

Here, the arrival rate attains a negative binomial distribution. Even though the arrival rate is *Negative Binomial* distributed, *Kendall's notation* depicts this as a general distribution (G). The service rate is depicted with a G, meaning the general distribution. The number of servers (z) in the system can be expressed by the number of judges in the system (eight for Family and eight for Youth) or in FTE (five for Family and six for Youth).

4.3 Conclusion

This chapter answers the question *How will the solution for F&Y be realised?* This is done by answering two sub-questions.

First, the question *How does the hearing length influence the idle time length between hearings?* is answered by the creation of a MILP model that minimises the idle time between hearings by seeking the optimal hearing length. The objective function of the MILP is to minimize the makespan. This model is applicable to different parameter values to provide results to answer the above question.

Second, *What are the parameters of F&Y's queuing system?* is answered. This is materialised with the identification of F&Y's queue as a $G/G/z$ queue. With this identification, we can make predictions about the utilisation, backlog, and waiting time under different experimental circumstances.

5 Results

Chapter 5 answers the question *To what extent does the solution contribute to the reduction of case backlog?* The MILP model created in *Chapter 4* is used to perform experiments to provide a solution to *Section 5.1*'s question. *Section 5.2* investigates the queuing aspect of the found solutions and the impact on backlog and waiting times.

5.1 Solution

We program the mathematical model in Python using the Gurobipy extension (Gurobi, 2025). The code is located in *Appendix C.1 Mathematical model*. Obtaining usable outcomes from the programmed model is not directly possible because the exact buffer size per case type is unknown (*Section 2.2*). To be able to answer this chapter's question, we conduct experiments.

5.1.1 Experimental design

We conduct two experiments. The first experiment is a sensitivity analysis, to measure the influence of the buffer size on the optimal hearing length (x_i^*). The second experiment is a robustness analysis, to measure the influence on the idle time of the optimal hearing length combinations, determined by the sensitivity analysis, under different buffer sizes. *Table 5* shows the experiment summary of the sensitivity analysis, *Table 6* of the robustness analysis. In these tables, the independent and dependent variables are set forth, together with the parameters used. In the robustness analysis, a control variable is present. These are the optimal hearing lengths as found in the sensitivity analysis.

The buffer lengths are examined between 0 and 1 hour per case. The choice for this lower and upper bound is based on *Section 2.2*. There, we determine that the maximum buffer size per hearing is 4 hours. Since there are currently 4 cases in a hearing (for most case types) $4\ buffer\ hours\ per\ hearing / 4\ cases = 1\ buffer\ hour\ per\ case$ is the maximum buffer size per case.

In the sensitivity analysis, results are examined under the buffer in steps of 0.25 hours. In the robustness analysis, steps of 0.1 hours are used to enhance preciseness of the results. In the robustness analysis, the hearing lengths are fixed according to the optimal hearing length combinations found in the sensitivity analysis.

Each experiment is executed for both the Family and Youth side separately since the model represents either a Family or a Youth judge and because of the earlier assumed mutual exclusiveness.

Sensitivity analysis	
	Variable
Independent	Buffer time per case (β_i)
Dependent	Optimal hearing length per case type (x_i^*)
	Parameter
Buffer size per case (β_i)	$\{0, 0.25, 0.5, 0.75, 1\}$
Hearing block length (θ)	4 hours
Big-M (M)	θ

Table 5: Experiment summary of sensitivity analysis.

Robustness analysis	
	Variable
Independent	Buffer time per case (β_i)
Dependent	Idle time
Control	Optimal hearing length per case type (x_i^*) (from sensitivity analysis)
	Parameter
Buffer size per case (β_i)	$\{0, 0.1, \dots, 0.9, 1\}$
Hearing block length (θ)	4 hours
Big-M (M)	θ
Hearing length (x_i)	Optimal hearing length per case type (x_i^*) (from sensitivity analysis)

Table 6: Experiment summary of robustness analysis.

5.1.1.1 Family side

For the Family side, we aggregate case types in groups because of their (i) similar characteristics in preparation, hearing, and finalisation times; (ii) because most Family judges are able to perform all these case types; (iii) and some of these case types are already scheduled together in hearings at F&Y; proving the possibility to schedule these case types together (*Section 2.1.3*). This saves computational time (because of symmetry prevention), without affecting the result as found in similar cases in the literature (*Section 3.1.2*). When aggregating, it will not be possible to give these case types unique parameter values concerning preparation, hearing, finalisation, and buffer time.

The Family case types' time parameter values in hours are shown by *Table 7*. Also shown here is the number of cases that need to be scheduled. This number is derived from the number of cases per case type that arrive monthly (4 weeks) and need to be scheduled on a hearing according to the probabilities explained by *Section 4.2.2*. The monthly arrival rate is divided by the total number of judges on the Family side (eight). Therefore, the *monthly judge workload* represents the average monthly workload of one judge.

Since the distribution between unscheduled cases and cancelled cases is unknown, we assume that the cases that do not have to come on a hearing also are not scheduled on a hearing;

Family		i	
Parameter	**Symbol**	**(G&O, ALI, ES, VOVO, BK1)**	**(KG, ESVD)**
Preparation time per case	p_i	1	2
Hearing time per case	h_i	1	2
Finalisation time per case	f_i	1	2
Monthly judge workload	ε_i	13	4
Hearing length set	x_i	$\{1,2,3,4\}$	$\{1,2,3,4\}$

Table 7: Parameter values for sensitivity analysis (Family).

5.1.1.2 Youth side

On the Youth side, the same calculations are applied as to the Family side concerning the *monthly judge workload* (*Table 8*). The VZ case type's hearing length is fixed at seven cases, similar to the current situation, because these cases take place outside of the court and thus are more efficient to group together (i.e. a driver is needed, and multiple locations have to be visited). Therefore, they are not taken into account when running the model. Leaving out the VZ case type does not affect the solution given the current preparation, hearing, and finalisation times.

Youth		*i*			
Parameter	**Symbol**	**EKJS**	**MKJS**	**VZ**	**OTS**
Preparation time per case	p_i	1.5	4	0.6	1
Hearing time per case	h_i	0.9	1.5	1.15	1
Finalisation time per case	f_i	0.8	1.5	0.6	1
Monthly judge workload	ε_i	2	1	11	14
Hearing length set	x_i	$\{1,2,3,4\}$	$\{1,2\}$	$\{7\}$	$\{1,2,3,4\}$

Table 8: Parameter values for sensitivity analysis (Youth).

5.1.2 Sensitivity Analysis

5.1.2.1 Family side

Figure 28 shows the influence of the buffer size on the optimal hearing length (x_i^*). Results are created by running the programmed model with the parameters as described by *Section 5.1.1*. For the KG and ESVD case type, no more than two cases are scheduled together; otherwise the hearing block break is overwritten (*Appendix A.2 Time prediction*). The model favours the hearing length that leaves no remainder when dividing the case input (ε_i) over the hearing processes. If there is a remainder from dividing the cases over the hearings processes, there is idle time in the hearings themselves, leading to suboptimal outcomes. For example, if case input (ε_i) is 10 for case type i, a x_i of 4 will by definition create idle time within the hearing: $\lceil 10\ cases/4\ cases\ per\ hearing \rceil = \lceil 2.5\ hearings \rceil \rightarrow 3\ hearings$. This results in 0.5 hearing of idle time withing the hearing.

According to the model, the optimal hearing length (x_i^*) is one case per hearing for the G&O, ALI, ES, VOVO, and BK1 case types except when the buffer is 0.25 hours per case. The optimal hearing length for the KG and ESVD case types is one case per hearing except when there is no buffer needed.

The sensitivity analysis yields three unique optimal hearing length combinations. They are indicated from here onwards with the following notation for the Family side:

$$(x^*_{G\&O+ALI+ES+VOVO+BK1} - x^*_{KG+ESVD})$$

The three unique hearing length combinations for the Family side are (1-2), (3-1), and (1-1).

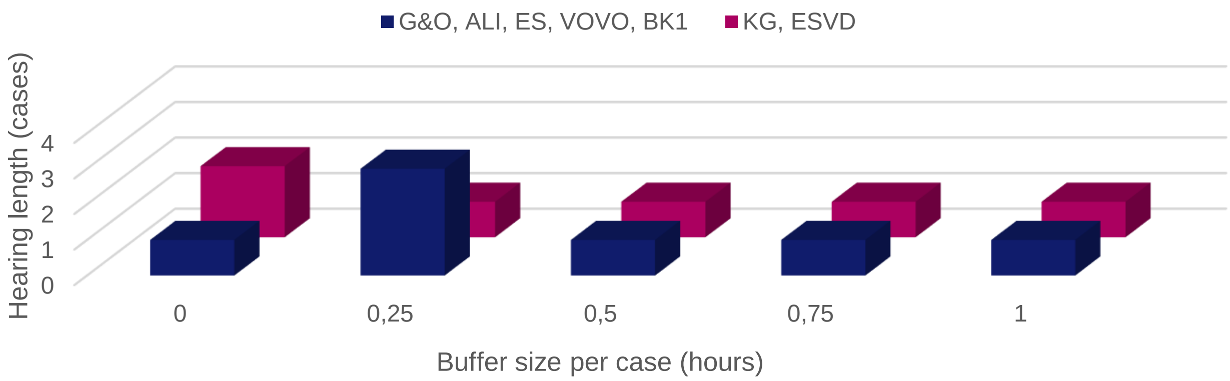


Figure 28: Influence of buffer size on optimal hearing length (Family).

Figure 29 shows the idle times that exist under the optimal hearing length combinations of the sensitivity analysis (*Figure 28*). The idle times are largest when the buffer is a fraction of an hour. This can be explained by the fact that a hearing process will probably be finished at a time ending on a decimal point. The next part of the hearing process can only start at an integer number; thus, idle time is created. The idle time decreases when the buffer size increases, because the sum of the idle time and the buffer is equal (*Section 1.2.2.4*).

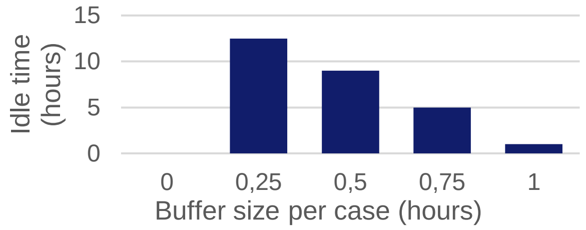


Figure 29: Idle time of optimal hearing length under differing buffer sizes (Family).

5.1.2.2 Youth side

Figure 30 shows the optimal hearing lengths (x_i^*) for the Youth case types under differing buffer lengths per case. As explained by *Section 5.1.1.2*, the VZ case type's hearing length is unchanged from the current situation (seven cases). The MKJS case type shows that the optimal hearing length is one case per hearing independent of the buffer length, because the arrival rate is one case; hence, a larger value of x_i will result in idle time within the hearing.

The optimal hearing length for the EKJS case type is two cases per hearing; independent of the buffer size. This is also related to its arrival rate (two cases). The OTS case type's optimal hearing length is one case per hearing, except under buffer lengths of 0.25 and 0.5 hours. There, the optimal hearing length is two cases per hour.

The Youth side's optimal hearing length combinations are described with the following notation:

$$(x_{EKJS}^* - x_{MKJS}^* - x_{VZ}^* - x_{OTS}^*)$$

The unique optimal hearing length combinations for the Youth side are (2-1-7-1) and (2-1-7-2).

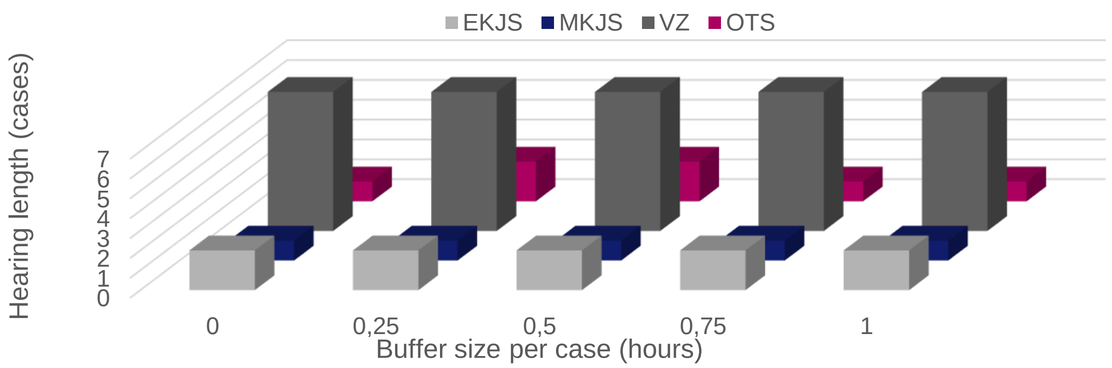


Figure 30: Influence of buffer size on optimal hearing length (Youth).

The idle times generated under the optimal hearing length combinations are shown by *Figure 31*. Contrary to the Family side, the buffer sizes do not show a stepwise change, indicating that the sum of the buffer and idle time is not equal in all situations. This is possible because there are more possible hearing orders due to the non-integer needed times per case; thus, creating more flexibility.

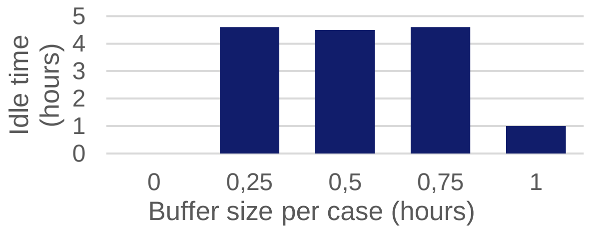


Figure 31: Idle time of optimal hearing length under differing buffer sizes (Youth).

5.1.3 Robustness analysis

The robustness analysis measures how the found optimal hearing length combinations in the sensitivity analysis influence the idle time under varying buffer sizes when these hearing length combinations are fixed.

5.1.3.1 Family side

Figure 32 shows the idle time evolution for the Family case types under different buffer sizes of the optimal hearing length combinations from the sensitivity analysis. These hearing length combinations are compared with each other and with the current situation of F&Y.

The hearing length combination currently attained by F&Y (4-2) yields significantly larger idle times than the found hearing length combinations. The case input is not divisible by 4 and 2 respectively, thus idle time exists within the hearings. For a buffer size between 0.1 and 0.3 hours, the optimal hearing length combination is (3-1). If the buffer size is between 0.3 and 0.5, both (1-2) and (1-1) yield equal idle times. When the buffer size is greater than 0.5 hours per case, the optimal hearing length combination is (1-1). It is remarkable that all found hearing length combinations yield better results than the current hearing length combination independent of the buffer size.

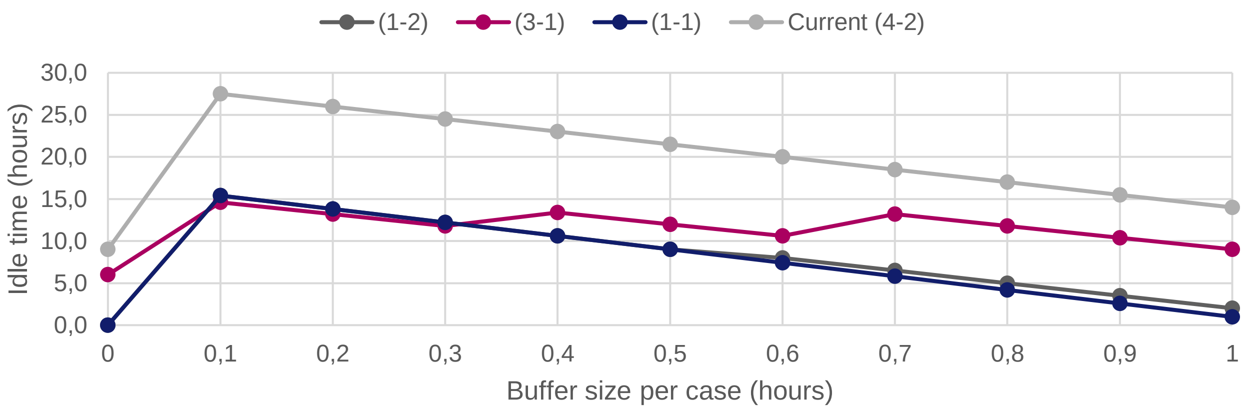


Figure 32: Influence of buffer size on idle time under fixed hearing lengths (Family).

5.1.3.2 Youth side

Figure 33 shows the idle time evolution under different buffer sizes for the Youth case types under the optimal hearing length combinations attained from the sensitivity analysis.

Contrary to the Family case type hearing length combinations, there is less of a linear trend visible in the development of idle time. The decimal input parameters concerning the preparation, hearing, and finishing times provide an explanation for this. These parameters offer more flexibility in the order of scheduling hearings, thus can reduce idle time more effectively.

When a buffer is needed between 0 and 0.6 hours per case, (2-1-7-2) is the optimal hearing length combination. When the buffer size is larger than 0.6 hours per case, there is no difference between (2-1-7-1) and (2-1-7-2) concerning idle time. The hearing length combination as attained currently by F&Y, (4-2-7-4), yields relatively similar idle times independent of the needed buffer size. Similar to the Family side, the current hearing length combination performs worse than all those found in the sensitivity analysis.

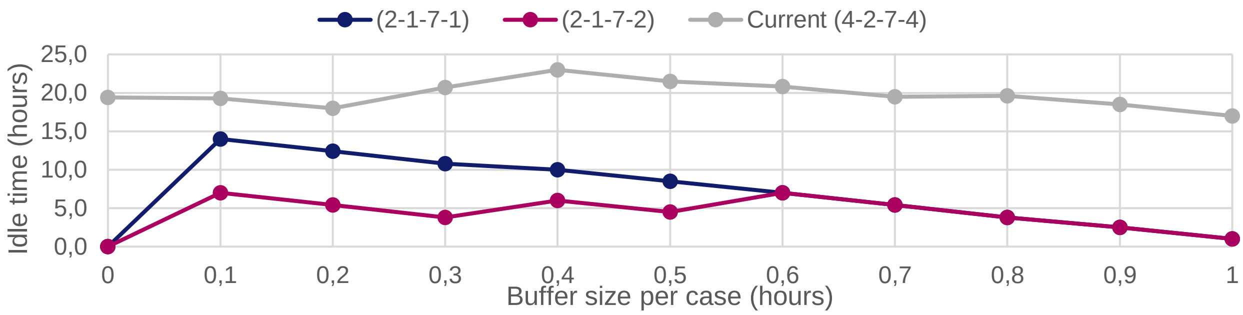


Figure 33: Influence of buffer size on idle time under fixed hearing lengths (Youth).

5.2 Effect on backlog

The queue parameters from *Section 4.2* are used to determine the effect of the found solutions. The *Allen-Cunneen* formula provides an approximation of the waiting time for a $G/G/z$ system – which is determined to be representative of F&Y's queue (Demanuele, 2015; Whitt, 1993). The *Allen-Cunneen* formula is build up as follows; where W_q is the mean waiting time in the queue, C_a^2 and C_s^2 are the variation coefficients of the interarrival time and service time respectively, ρ the utilisation, c the number of servers, and μ the service rate (Pesant et al., 2024):

$$W_q \approx \frac{C_a^2+C_s^2}{2} * \frac{\rho^{\sqrt{2(c+1)}-1}}{c*(1-p)} * \frac{1}{\mu}$$

To calculate the mean waiting time of the entire system (W), we use the following formula:

$$W = W_q + \frac{1}{\mu}$$

The *Allen-Cunneen* formula solely provides an approximation of the mean waiting time in the queue and in the system. A precise prediction cannot be made because of the queuing system ($G/G/z$).

Section 5.2.1 approximates the backlog that should theoretically exist at F&Y Haarlem currently. *Section 5.2.2* approximates the backlog development under the optimal hearing length combinations that are found in the sensitivity analysis and explored in the robustness analysis.

5.2.1 Theoretical backlog

Table 9 shows F&Y's theoretical values of the parameters needed for the *Allen-Cunneen* formula – grouped separately for the Family and Youth side. Since the model represents a fulltime working judge, the number of c is applied in Full Time Equivalent (FTE) – not in number of judges. The values of variance of C_a^2 and C_s^2 are calculated based on the arrivals and the *theoretical judge capacity* respectively with the following formula, where s^2 is the sample variance, n the sample size, and x the mean of the sample (Meijer, 2022):

$$C_a^2, C_s^2 = s^2 = \frac{1}{n-1} * \sum_{i=1}^{n}(x_i - x)^2$$

	Family	Youth
C_a^2	0.0785	0.0634
C_s^2	0.0681	0.0191
ρ	0.6305	0.7804
c	5	6
μ	52	73

Table 9: F&Y's theoretical values of parameters for Allen-Cunneen formula.

The *Allen-Cunneen* formula provides an approximation for the waiting time in the queues (W_q); with a value of close to zero weeks for both the Family and Youth side. This results in mean waiting times of the entire system of 0.0193 and 0.0139 weeks for the Family and Youth side, respectively. This applies to the theoretical situation, where no idle time is present.

Little's law is used to determine the number of customers in the system. We deem this formula valid to use because it is also applied in similar situations as described by *Section 3.2.1*. Here, L represents the average number of customers – cases – in the system, λ the arrival rate, and W the average waiting time of the customers in the system (Stidham, 1974):

$$L = \lambda W$$

Applying *Little's law* to the theoretical scenario provides the numbers of customers in the system (L) of 0.6384 and 0.7933 on average for the Family and Youth sides, respectively.

These numbers differ significantly from the observed values of case backlog (*Section 2.3.3*), showing again that the theoretical values do not correspond with reality. Based on the available data and information, this confirms that the *theoretical judge capacity* is overestimated (*Section 1.2.2.2*).

5.2.2 Solution backlog

The outcomes of the robustness analysis alter the service rate because of increased needed time for cases due to the buffer and idle time. The waiting time and backlog size are in turn influenced by the service rate.

5.2.2.1 Family side

Figure 34 shows the service rate under fixed hearing length combinations together with a theoretical service rate for the Family side. This theoretical service rate is determined similarly as by *Section 4.2.2*. Here, the needed buffer size is added on top of the average case duration; thus, influencing the service rate. In the theoretical situation, no idle time is present.

Under the hearing length combinations, the service rate attains a near equal value independent of the buffer size. This for the fact that the sum of the buffer and idle time remains equal. An exception to this is the (3-1) combination since its service rate changes due to different possible hearing scheduling orders.

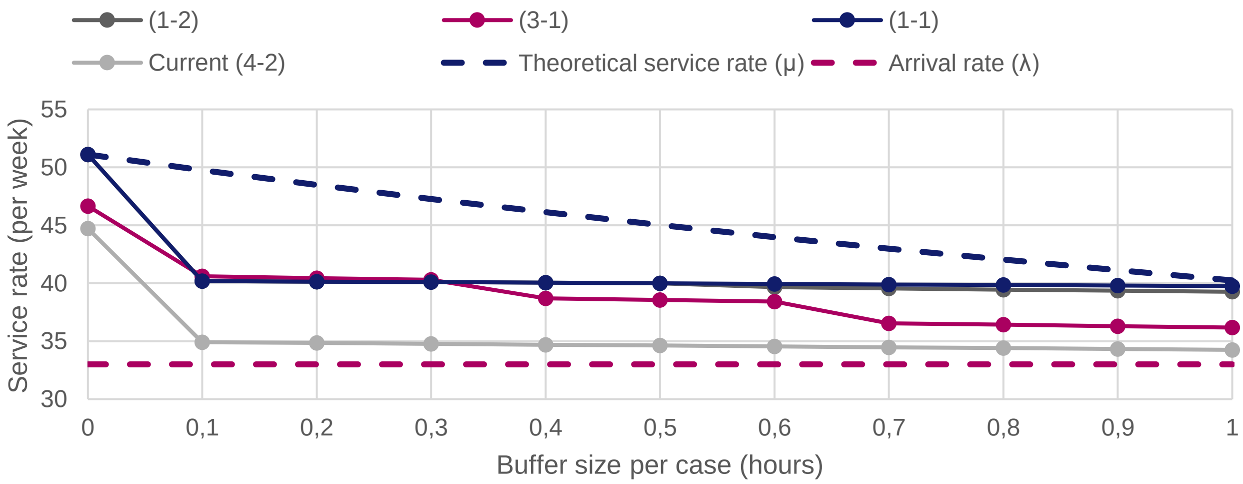


Figure 34: Service rate of hearing length combinations and theoretical case (Family).

Figure 35 shows the development of the average queue length (L) – in cases – under differing buffer sizes for the fixed hearing length combinations of the Family side. Also here, the theoretical growth in queue length is shown. All found hearing combinations perform worse than the theoretical situation but perform significantly better than the current situation.

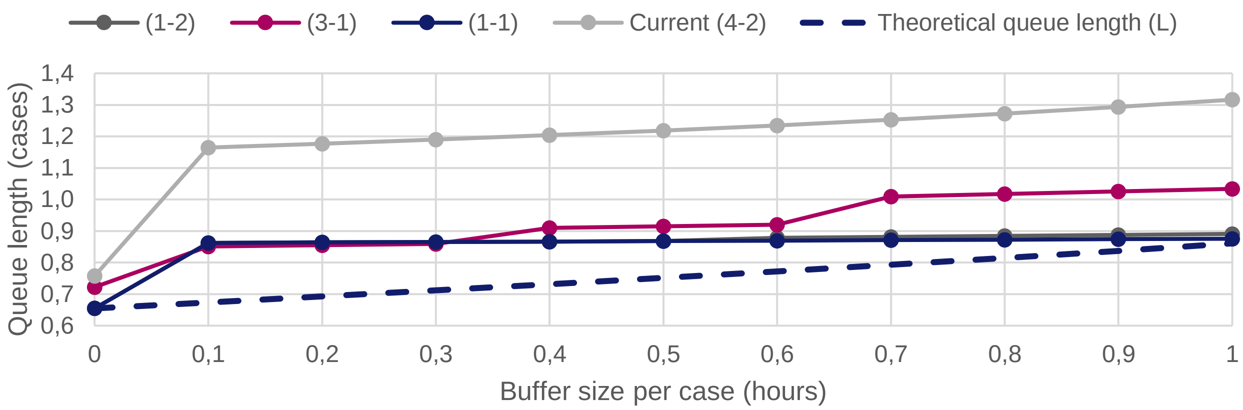


Figure 35: Queue length (L) under fixed hearing length combinations and buffer (Family).

For the Family side, in general, the optimal hearing length combination is (1-1) for the used parameter values. Even though (1-2) provides a roughly similar reduction in waiting time and queue length, (1-1) yields better results the larger the buffer is. (3-1) yields the greatest reduction under smaller buffer sizes; however, performs significantly worse the larger the buffer is (*Figure 36*).

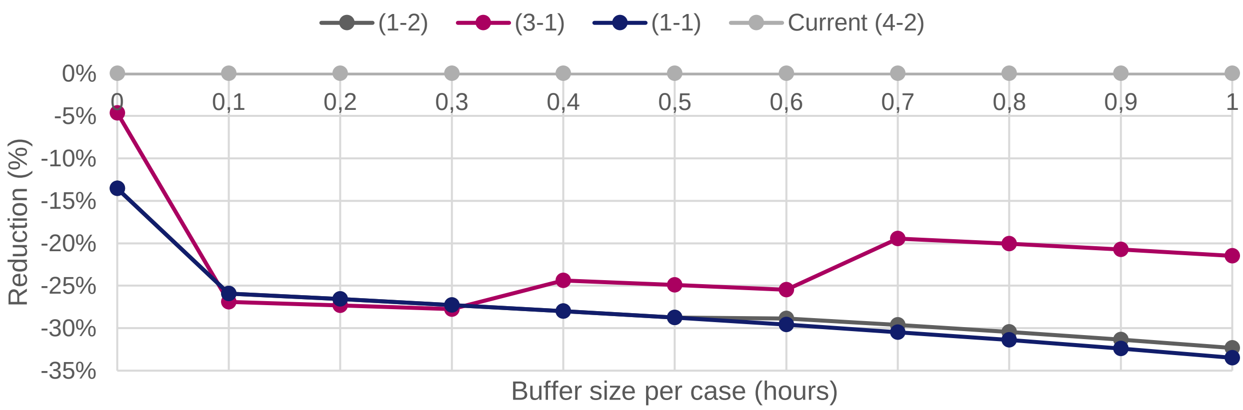


Figure 36: Waiting time (W) and queue size (L) reduction compared to current situation (Family).

5.2.2.2 Youth side

Figure 37 shows the service rate development of the optimal hearing length combinations under varying buffer sizes for the Youth side. Different from the Family side is that the service rate is smaller than the arrival rate for some hearing length combinations under some buffer sizes. This results in a utilisation higher than one, meaning that the system explodes *ipso facto*. Under the current situation, the utilisation is always greater than one.

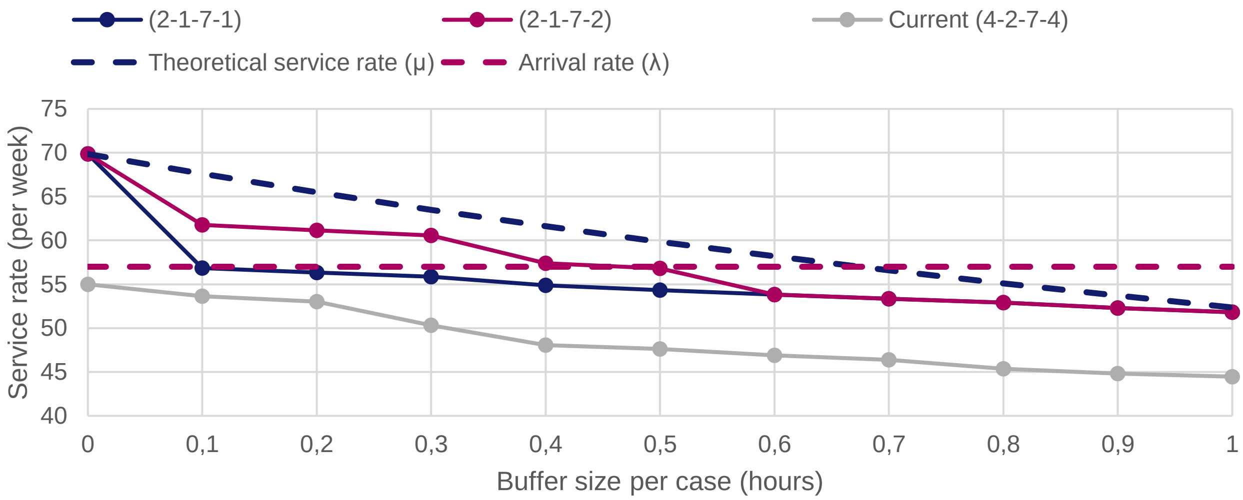


Figure 37: Service rate of hearing length combinations and theoretical case (Youth).

Figure 38 shows the queue length (L) evolution in cases. It can be seen that the system explodes at a certain buffer size because of the utilisation being greater than one. When the utilisation is greater than one, the *Allen-Cunneen* formula is not applicable. Therefore, not all hearing length combinations are shown, including the queue length under the current situation.

Under the current situation it is not possible to keep up with demand. This explains why there is growing backlog. The only hearing length combination that is workable is (2-1-7-2). However, only when the buffer is between 0 and 0.4 hours per case.

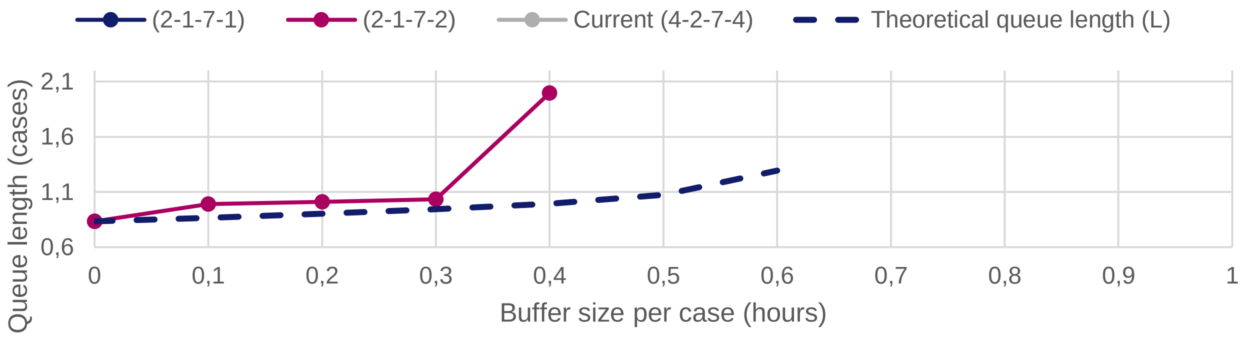


Figure 38: Queue length (L) under fixed hearing length combinations and buffer (Youth).

Because there is no percentual improvement of the current hearing length combination, *Figure 39* shows the utilisation rates of the hearing length combinations instead. Here, we see that the utilisation is greater than one under most experiment circumstances.

Based on the high utilisation rates, it seems counterintuitive that there is no backlog of Youth case types at F&Y Haarlem. A theory to explain this is that the Youth case types have stricter deadlines for verdicts, thus have a higher priority to be scheduled in time. Since the assumed mutual exclusivity of the Family and Youth sides is not as strict in reality, there are Family judges that can perform Youth case types (i.e. VZ and OTS); thereby, increasing the service rate of the Youth side. Family judges that are put to work on Youth case types result in undercapacity at the Family side, whose case types have "less" priority. Thus, a utilisation larger than one exists at the Family side and consequently growing backlog as shown in the backlog numbers by *Section 2.3.3*.

Shifting capacity in a way that the utilisation rate is lower than one – for both the Family and the Youth side – is possible according to the maximum capacity per case type (*Section 2.1.2.3*). However, determining exactly how many FTE should be transferred from one side to the other is difficult to say because of the assumed mutual exclusivity of the Family and Youth side, and the different case specialisations per judge.

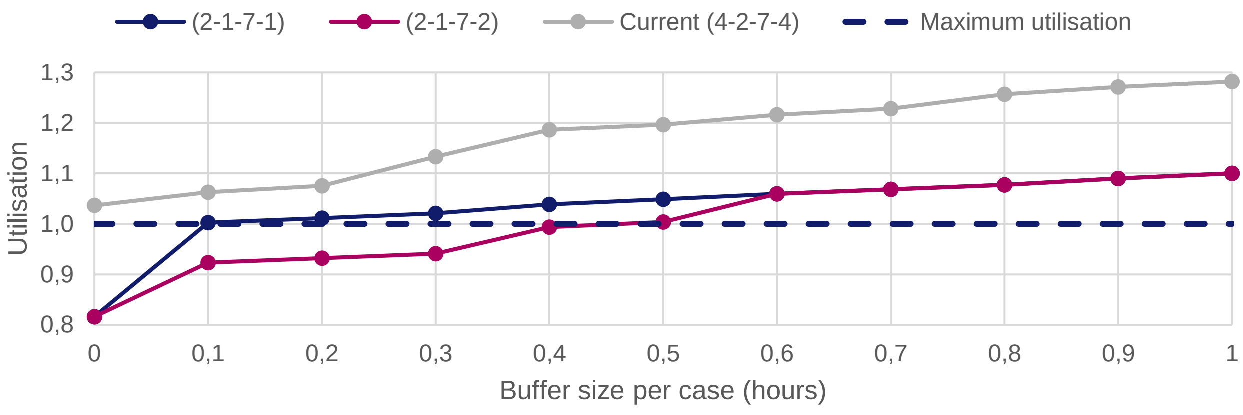


Figure 39: Utilisation of the optimal and current hearing length combinations (Youth).

5.3 Conclusion

This chapter answers the following question: *To what extent does the solution contribute to the reduction of case backlog?*

By performing a sensitivity analysis, a robustness analysis, and a queuing analysis of those outcomes; we find that the current hearing length combination is performing far under the theoretical performance. Both in terms of service rate as well as in queue length. This shows the discrepancy between the theoretical and realised situation; proving the existence of idle time.

The found optimal hearing lengths in the sensitivity analysis perform better in the robustness analysis, under all circumstances, than the current hearing lengths. This shows the possibility to reduce idle time by adjusting the hearing lengths.

For the Family case types, backlog and waiting time reduction is between 13.6 and 33.5 percentage points compared to the current situation. For the Youth side, the reduction is not measurable compared to the current situation for the fact that the current situation yields a utilisation larger than one.

6 Discussion

We made numerous assumptions as foundation for this report. These assumptions are needed to fill gaps created by uncertainty of unknown information. Therefore, a discussion is needed to critically reflect on the context of this research. First, *Section 6.1* sets forth the contributions of this research. Then, *Section 6.2* discusses the limitations of the research together with their effect on the overall reliability and validity.

6.1 Contributions

The results of this research contribute to both the Family and Youth section of Court of Law Noord-Holland as well as to the scientific community in general.

6.1.1 Contribution to Family & Youth Court of Law Noord-Holland

A direct contribution to F&Y is the ability to make the size of idle time insightful. Furthermore, the solution offers flexibility in the hearing schedule and makes it possible to reduce idle time by finding the optimal hearing length per case type.

Finding optimal hearing lengths is not only possible under the current *predicted needed times* for case types, but also when the *predicted needed times* for case types are recalculated. Then, the model will also be able to provide solutions that minimise the idle time.

We propose procedures to predict backlog and waiting time development under different parameter values of needed times. To the knowledge of the author, there have not been made approximations of the waiting time so far.

This research focusses on the Haarlem court because the hearing schedules are currently made separately for the Alkmaar and Haarlem courts. The solution is flexible enough to be also applicable to the Alkmaar court when the necessary data is available.

The features of a successful policy that Bray et al. (2016) point out – as discussed in *Section 3.2.1* – are fulfilled: The policy proposed is beneficial for F&Y; it also preserves judicial autonomy because judges can still plan their day within the preparation and finalisation blocks; it does not increase the workload of judges, it only reduces the idle time and prepares the hearing schedule to be updated with new and realistic *predicted needed times* for hearings; and finally, it is straightforward to understand and implement the solution by the schedulers since it requires basic input data and does not change the structure of the way the hearing schedule is made.

6.1.2 Contribution to science

This research contributes to the scientific community in two ways:

First, by applying linear programming to the judicial system. Linear programming is little applied to the judicial system currently. This research adds knowledge about its applications and distinctions compared to other fields in the service sector. It also looks at the hearing structure instead of only at the hearing scheduling; opening up new possibilities to improve and optimise hearing schedules.

Second, the queuing system is identified. Contrary to other judicial systems, which are mostly identified as $M/M/z$ according to the literature found by *Section 3.2.1*, F&Y is identified to be a $G/G/z$ system. This yields new insights in approximating backlog and waiting times for similar systems.

6.2 Limitations

We encountered many limitations during the creation of this report. A wide-reaching limitation is time related; merely ten weeks are available for the execution of the research. This limits the depth and extensiveness of the research. Nevertheless, effort is undertaken to maximize the impact, reliability, and validity of the results. We explain more limitations in the sections below, sorted per topic. The effect of these limitations on the reliability and validity of the results is also elaborated on.

6.2.1 Model

One of the limitations encountered with the development of the MILP model is that it represents a fulltime judge that has a primary hour rate of one; meaning that the judge can spend all hours in a workday on hearings and cases, this is not representative for all F&Y judges. Also, the model does not consider preferences of judges related to case types and preferred hearing times.

The model can only incorporate case types that have a maximum duration of one day part (four hours). Therefore, currently, optimal hearing lengths of case types such as VZ cannot be determined by the model without limiting the maximum hearing duration to four hours, since these case type’s hearings take an entire day (eight hours). Additionally, case types with a hearing length of more than four cases are not incorporated. Although this is possible in the model, it is not done due to the fact that most case types are scheduled with a hearing length of four cases per hearing.

The model does not consider the deadlines connected to cases, hearings, and verdicts. This can pose problems if there is a weekend or holiday between hearing parts as explained by *Section 2.1.2.2*.

In the execution of the model, we see that the optimal hearing length (x_i^*) is strongly dependent on the number of cases to schedule because of idle time within hearing processes. When a remainder exists after dividing the cases over the hearing processes, there is empty room in a hearing, and thus idle time within the hearing process. This leads to suboptimal results. In reality there is continuity of case scheduling, thus this bias of x_i based on idle time within hearing processes – instead of between – is less present.

The model provides reliable results. However, the validity of the outcomes is negatively influenced by the inability to incorporate all judge and case type parameters.

6.2.2 Data

A major limitation of the solutions proposed lies within the availability, consistency, and completeness of case input, output, and backlog data.

Case input data is incomplete for the EKJS and MKJS case types, possibly providing false assumptions about consistency of case input over the years. This leads to an inaccurate assumed arrival rate.

For the arrival rate used in the sensitivity and robustness analysis, we assume that all cases that do not come on a hearing are also unscheduled. This is unrealistic since it does not take into account the cases that are scheduled but cancelled before the hearing takes place – requiring time to prepare, resulting in a decreased service rate. However, the extent of the incorrectness of this assumption is unknown because the distribution between unscheduled and cancelled cases is unknown.

Case output data is unavailable. Therefore, we base the service rate on the *theoretical judge capacity*, which does not correspond with the realised service rate as we showed in multiple sections of this report. Truthful solutions can be created when this data is available. Then, the *realised judge capacity* can be taken into account to create valid results.

The arrival rate is determined based on data from 2019 up and until 2024 while the service rate takes into account the current judge capacity (1 May 2025). This can provide a false picture of the relation between the arrivals and the service capacity – the utilisation.

The extent of the chosen action problem of this report is unknown for the fact that the exact size of case backlog is unknown. This makes it difficult to determine the severity of the problem and the associated waiting times for justice seekers. Furthermore, the evolution of case backlog over time is unknown.

The lack of the abovementioned data harms the validity of the created solutions. However, due to the reliability of the model, the current incomplete data does not pose a problem when the input, output, and backlog numbers are established. The model will yield reliable and valid results once they are realistically determined and used as input for the model.

6.2.3 Experimentation

For the sensitivity, robustness, and queuing analysis of this report, we aggregated case types with similar parameters values (i.e. preparation, hearing, and finalisation times). This decreases runtime of the model, while still creating valid results. On the other hand, it limits the ability to give these case types unique parameter values concerning the buffer size.

Continuing on the buffer size, in the experiments we applied the same independent variable (buffer size) to all case types simultaneously. This is highly unlikely to be the case in reality, but it is applied to be able to provide a clear visualisation of the relation between the buffer size and the idle time. To increase reliability, unique buffer sizes should be attributed per case type.

The used formula for the queuing analysis – the *Allen-Cunneen formula* – only provides an approximation of the waiting times in the system. Hence, this approximation is imprecise and thus unreliable. Because of the queue type of F&Y ($G/G/z$), it is difficult to make exact estimations of the developments in waiting times and case backlog.

6.2.4 Assumptions

We made assumptions to be able to provide results and create solutions. The main assumptions are discussed here.

A core assumption, serving as a foundation of this report, is that the Family and Youth side are mutually exclusive. It is known that this is untrue since Family judges can perform Youth case types (e.g. VZ and OTS) and vice versa. However, we need this assumption to determine the *theoretical judge capacity* to in turn determine the service rate of F&Y. This approach is a possible unreliable choice. Reliability can be increased when the output data is known and used to determine the *realised judge capacity* as mentioned by *Section 6.2.2*.

Concerning the judge specialisations, we assume that the Family judges are able to perform all Family case types and that all Youth judges can perform all Youth case types. In reality, judges have different specialisations and are not capable to perform all case types of their respective side. We also made this assumption to be able to determine the service rate. Solving this unreliability is done similarly as proposed in the paragraph above.

We assume that the maximum buffer size is one hour per case. Whether one hour buffer per case is enough is unknown, but probably enough as proposed by *Section 2.2.4*. In the case that this limit is not reaching the needed buffer size, it can be easily adjusted in the model. Experiments can be altered to a new buffer size range, showing the reliability and flexibility of both the model and the experiments.

The assumption that the resource scarcity is located at the judges at all times is true when looking at averages of a year. However, it might be that there is seasonal change in capacity between judges and clerks (e.g. both parties take holidays at different times of year).

We made a similar assumption for the case input. Consistent case input throughout the year, and over the years, is assumed. This may be correct when looking at the total case input, but perhaps not when looking per case type. Therefore, validity of the results may be harmed because of this significant assumption.

On a hearing, cases are given predicted needed times based on the contents of the case (i.e. the size of the contents, the number of interested parties involved). Here, we assume that cases' predicted times are the *predicted needed times* for a hearing divided by the number of cases on that hearing. Thereby, generalising all cases of a type. Further validity can be reached when these case specific needed times are considered for the generation of solutions.

A significant threat to validity is the inability to measure the effect of the solution on the backlog and waiting times with certainty. Furthermore, more abstract variables such as workload and quality of verdicts are even more difficult to determine.

7 Recommendations & Conclusion

This chapter proposes recommendations for F&Y in *Section 7.1* – related to the hearing length and the hearing scheduling process. Possible future research is determined by *Section 7.2*. Finally, *Section 7.3* answers the research question of this research – *How can section Family & Youth at the Court of Law Noord-Holland reduce case backlog by adjusting the number of cases in hearings?*.

7.1 Recommendations

The findings of this report provide a foundation for recommendations concerning F&Y's hearing scheduling. *Section 7.1.1* provides recommendations concerning the hearing length of F&Y's case types. *Section 7.1.2* provides recommendations concerning the broader scheduling process.

7.1.1 Hearing length

Under the current arrival rate and time prediction parameter values, the recommendation is to change the hearing length as shown by *Table 10*. However, these hearing lengths may not be optimal in reality under the parameter values used. Therefore, we do not necessarily recommend to implement these proposed hearing lengths, rather to take a critical look at the data collection and administration of case input, output, and backlog. Once these values are structured and insightful, they can be implemented in the model to produce substantiated decisions concerning a valid optimal hearing length. Also, good approximations of the backlog and waiting time development can be made once the necessary input data is available. A dashboard is a feasible tool to provide this much needed overview of data.

	EKJS	MKJS	ESVD	BK1	ES	ALI	VOVO	KG	G&O	VZ	OTS
Current x_i	4	2	2	4.5	4	4	4.5	2	4	7	4
Recommended x_i	2	1	1	1	1	1	1	1	1	7	2

Table 10: Recommended hearing lengths.

7.1.2 Hearing scheduling process

Concerning the broader hearing scheduling process, we recommend to schedule cases directly instead of scheduling hearing processes and then filling these with cases. This removes idle time within hearings. Furthermore, more flexibility is created in the schedule, providing opportunity to reduce idle time between hearings. This way of scheduling is already applied at section *commercial law* of Court of Law Noord-Holland, among others.

Judges and clerks should be scheduled more coherent, since both the judges and clerks are dependent on each other in every stage of their intertwined hearing process (*Figure 40*). The separate hearing scheduling of the two parties can be a cause of idle time for the fact that both parties have to wait on each other for starting their hearing parts. Coherency can be created by scheduling the hearing parts of both the judges and clerks together when making the schedule.

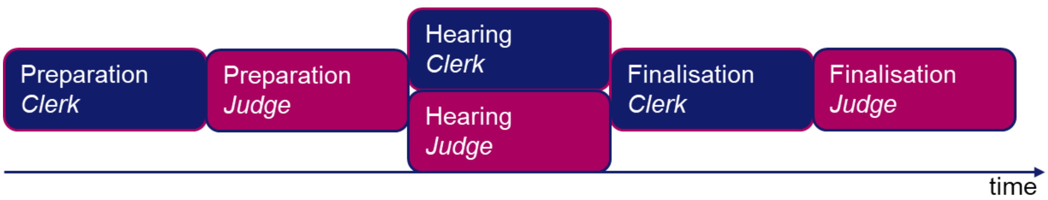


Figure 40: Judge and clerk dependency in the hearing process.

To provide more flexibility in the hearing schedule – to reduce idle time creation, hearings could start at half hours. Currently, hearings are scheduled at whole hours. When they can also be scheduled at half hours, it will provide more flexibility and opportunities to schedule hearings. On a regular day there are currently eight moments to start a hearing. When half hours are used, there are sixteen possible start moments.

7.2 Future research

The research performed at F&Y Court of Law Noord-Holland is far from finished. It is merely a starting point to improve the hearing schedule. In future research, several points have to, and can, be addressed.

The discussed threats to reliability and validity, as discussed by *Chapter 6*, need to be addressed and resolved to produce workable and valid solutions for F&Y concerning the hearing lengths.

In future research, other topics can be explored such as the effect of different buffer lengths per case type on the optimal hearing lengths. Also, the effect of case aggregation on the needed buffer length needs to be explored since buffer lengths needed per case are not independent once cases are aggregated. The hearing length influences the variability of the *real needed time* for hearings (Hans et al., 2008).

The way of scheduling at other sections at Court of Law Noord-Holland, and at different courts, should be explored because of different approaches to hearing and case scheduling among the decentralised judicial organisations. An example of this is the earlier mentioned direct case scheduling of the *commercial law* section of Court of Law Noord-Holland.

A possible sector to look into is the aviation industry. This because of possibly overbooking hearings with cases based on the probability of a case cancelling before a hearing. Thereby reducing idle time if a hearing can still be filled after a case is cancelled.

The need for the hearing block break at 13:00 should be investigated. Once it can be dropped, there will be more flexibility in the schedule, which further reduces idle time.

7.3 Conclusion

We addressed the chosen action problem – *the existence of too large case backlog and too long waiting times for justice seekers*. For that, we attempted to solve the identified core problem – *The hearing length forces the scheduling of idle time between hearings if a buffer is scheduled* – by answering the research question: *How can section Family & Youth at the Court of Law Noord-Holland reduce case backlog by adjusting the number of cases in hearings?*

After identifying the hearing scheduling process of F&Y and searching relevant literature related to linear programming and queuing theory in the judicial system and the broader service sector, we proposed a MILP to minimise the idle time scheduled between hearings. Furthermore, the queue parameters of F&Y are identified as a $G/G/z$ queue.

Applying the available data to the model yields results that we evaluated with a sensitivity, robustness, and queuing analysis. We find that under the current assumptions and available data, the hearing length should be altered to minimise idle time between hearings. The current hearing lengths attained by F&Y perform worse in the robustness analysis compared to the optimal hearing lengths, as found in the sensitivity analysis. Furthermore, the current hearing lengths perform worse than the theoretical situation; showing both the existence of idle time, and that other hearing lengths provide less idle time within the schedule.

To reduce case backlog, and corresponding waiting times, we recommend F&Y to improve the organisation and administration of data to be able to realistically determine the extent to which the hearing lengths need to be adjusted. Thereafter, the reduction of idle time, backlog, and waiting time of those solutions can be determined if the queue parameters of F&Y are known.

References

Babaei, H., Karimpour, J., & Hadidi, A. (2015). A survey of approaches for university course timetabling problem. *Computers & Industrial Engineering*, *86*, 43–59. https://doi.org/10.1016/j.cie.2014.11.010

Bakshi, N., Kim, J., & Randhawa, R. S. (2025). Service Operations for Justice-on-Time: A Data-Driven Queueing Approach. *Manufacturing and Service Operations Management*, *27*(1), 305–321. Scopus. https://doi.org/10.1287/msom.2023.0530

Bixby, R. E. (2012). A brief history of linear and mixed-integer programming computation. *International Symposium on Mathematical Programming*, *21*. https://doi.org/10.4171/DMS/6/16

Boccia, M., Bruno, G., & Sterle, C. (2013). *A MILP formulation for a batch scheduling problem on parallel machines in the aircraft industry*. 2013 5th International Conference on Modeling, Simulation and Applied Optimization, ICMSAO 2013. Scopus. https://doi.org/10.1109/ICMSAO.2013.6552650

Brahimi, M., & Worthington, D. J. (1991). Queueing Models for Out-Patient Appointment Systems—A Case Study. *The Journal of the Operational Research Society*, *42*(9), 733–746. https://doi.org/10.2307/2583656

Bray, R. L., Coviello, D., Ichino, A., & Persico, N. (2016). Multitasking, multiarmed bandits, and the Italian judiciary. *Manufacturing and Service Operations Management*, *18*(4), 545–558. Scopus. https://doi.org/10.1287/msom.2016.0586

Brooks, J. P. (2012). The Court of Appeals of Virginia Uses Integer Programming and Cloud Computing to Schedule Sessions. *Interfaces*, *42*(6), 544–553. https://doi.org/10.1287/inte.1110.0598

Chavannes, M. (2019, November 28). *Rechters en hun medewerkers hebben het veel te druk. Maar dat is niet het ergste: De rechtspraak lijdt eronder*. De Correspondent. https://decorrespondent.nl/10751/rechters-en-hun-medewerkers-hebben-het-veel-te-druk-maar-dat-is-niet-het-ergste-de-rechtspraak-lijdt-eronder/a4d69b93-b4ed-09e4-00e8-3bc5db2c864a

Chen, D.-S., Batson, R. G., & Dang, Y. (2010). *Applied Integer Programming: Modeling and Solution*. Wiley. https://www.wiley.com/en-us/Applied+Integer+Programming%3A+Modeling+and+Solution-p-9780470373064

Cho, K. W., Kim, S. M., Chae, Y. M., & Song, Y. U. (2017). Application of Queueing Theory to the Analysis of Changes in Outpatients’ Waiting Times in Hospitals Introducing EMR. *Healthcare Informatics Research*, *23*(1), 35–42. https://doi.org/10.4258/hir.2017.23.1.35

de Rechtspraak. (2019, November 6). *Jaardocumenten Raad voor de rechtspraak*. Het tijdbestedingsonderzoek in relatie tot de productie gerelateerde bijdrage voor het primair proces van de rechtspraak. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten/Paginas/default.aspx

de Rechtspraak. (2020, January 1). *Wet zorg en dwang (Wzd)*. Wet zord en dwang (Wzd). https://www.rechtspraak.nl/Onderwerpen/wzd/Paginas/default.aspx

de Rechtspraak. (2021a). *Mission, vision and agenda of The Netherlands Judiciary*. 1–4. https://www.rechtspraak.nl/SiteCollectionDocuments/mission-vision-and-agenda-of-the-judiciary.pdf de Rechtspraak. (2021b, January 19). *Professionele standaarden van rechters*. Professionele standaarden van rechters. https://www.rechtspraak.nl/Organisatie-en-contact/Rechtspraak-in-Nederland/Rechters/Paginas/De-professionele-standaarden-van-de-rechters.aspx

de Rechtspraak. (2023). *Rechtspraak in cijfers*. Rechtspraak in cijfers. https://www.rechtspraak.nl/Organisatie-en-contact/Rechtspraak-in-Nederland/Rechtspraak-in-cijfers/Paginas/default.aspx

de Rechtspraak. (2024a). *Rechtspraak Jaarverslag 2023*. https://www.rechtspraak.nl/SiteCollectionDocuments/Jaarverslag%20Rechtspraak%202023.pdf

de Rechtspraak. (2024b, October 30). Zaaksverdelingsreglement rechtbank Noord-Holland. *Staatscourant van het Koninkrijk der Nederlanden*. https://zoek.officielebekendmakingen.nl/stcrt-2024-35131.html

de Rechtspraak. (2025a, January 1). *Dutch procedural law*. Dutch Procedural Law. https://www.rechtspraak.nl/English/NCC/Pages/rules.aspx

de Rechtspraak. (2025b, January 1). *Familie- en jeugdrecht—Reglementen, procedures en formulieren*. Familie- en jeugdrecht - reglementen, procedures en formulieren. https://www.rechtspraak.nl/Voor-advocaten-en-juristen/Reglementen-procedures-en-formulieren/Civiel/Familie-en-jeugdrecht/Paginas/default.aspx

de Rechtspraak. (2025c, April 7). *Korte uitleg over 'bodemprocedure'*. https://www.rechtspraak.nl/juridische-begrippen/Paginas/bodemprocedure.aspx

Demanuele, T. (2015). *Analysis on queueing systems having general inter-arrival time and service time distributions* [University of Malta]. https://www.um.edu.mt/library/oar/handle/123456789/93805

Dimopoulou, M., & Miliotis, P. (2001). Implementation of a university course and examination timetabling system. *European Journal of Operational Research*, *130*(1), 202–213. https://doi.org/10.1016/S0377-2217(00)00052-7

Duijneveldt, I., van, Wijga, P., & Reisen, K., van. (2016). Naar een vitale organisatie: Een duurzaam antwoord op werkdruk binnen de Rechtspraak. *Research Memoranda*, *12*(2), 95.

European Court of Human Rights. (2013). European Convention on Human Rights. *Council of Europe*.

Ferdinandes, M. G. R. U. K., Lanel, G. H. J., & Samrarakoon, M. A. S. C. (2017). A Queuing Model to Optimize the Performance of Surgical Units. *International Journal of Advanced Engineering Research and Science*, *4*(5), 237170.

Fowler, J. W., & Mönch, L. (2022). A survey of scheduling with parallel batch (p-batch) processing. *European Journal of Operational Research*, *298*(1), 1–24. https://doi.org/10.1016/j.ejor.2021.06.012

Fowles, J. (1974). On chronocentrism. *Futures*, *6*(1), 65–68. https://doi.org/10.1016/0016-3287(74)90008-1

Gibbons, J. D. (2003). *Nonparametric statistical inference* (4th ed., rev.expanded).

Graham, R. L., Lawler, E. L., Lenstra, J. K., & Kan, A. H. G. R. (1979). Optimization and Approximation in Deterministic Sequencing and Scheduling: A Survey. In P. L. Hammer, E. L. Johnson, & B. H. Korte (Eds.), *Annals of Discrete Mathematics* (Vol. 5, pp. 287–326). Elsevier. https://doi.org/10.1016/S0167-5060(08)70356-X

Gupta, M., & Bolia, N. B. (2024). Redistribution of judicial resources for improved performance. *Annals of Operations Research*, *342*(3), 2147–2168.

Gurobi. (2025, May 13). *Gurobi Optimizer Reference Manual*. https://docs.gurobi.com/projects/optimizer/en/current/reference/python.html

Hans, E., Wullink, G., van Houdenhoven, M., & Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, *185*(3), 1038–1050. https://doi.org/10.1016/j.ejor.2006.08.022

Heerkens, H., & Winden, A. van. (2017). *Solving Managerial Problems Systematically*. Noordhoff Uitgevers. https://research.utwente.nl/en/publications/solving-managerial-problems-systematically

Huang, J. (2025). Minimizing makespan for mixed batch scheduling with identical machines and unequal ready times. *Scientific Reports*, *15*(1). Scopus. https://doi.org/10.1038/s41598-025-89698-3

Hulshof, P. J. H., Kortbeek, N., Boucherie, R. J., Hans, E. W., & Bakker, P. J. M. (2012). Taxonomic classification of planning decisions in health care: A structured review of the state of the art in OR/MS. *Health Systems*, *1*(2), 129–175. https://doi.org/10.1057/hs.2012.18

Hwang, F. J., Kovalyov, M. Y., & Lin, B. M. T. (2012). Total completion time minimization in two-machine flow shop scheduling problems with a fixed job sequence. *Discrete Optimization*, *9*(1), 29–39. Scopus. https://doi.org/10.1016/j.disopt.2011.11.001

İnce, N., Deliktaş, D., & Hakan Selvi, İ. (2024). A comprehensive literature review of the flowshop group scheduling problems: Systematic and bibliometric reviews. *International Journal of Production Research*, *62*(12), 4565–4594. Scopus. https://doi.org/10.1080/00207543.2023.2263577

Jennings, J. B. (1971). *A Theory of Court Scheduling*. RAND Corporation. https://www.rand.org/pubs/papers/P4732.html

Ji, B., Zhang, S., Yu, S. S., & Zhang, B. (2023). Mathematical Modeling and A Novel Heuristic Method for Flexible Job-Shop Batch Scheduling Problem with Incompatible Jobs. *Sustainability*, *15*(3), Article 3. https://doi.org/10.3390/su15031954

Jinsong, B., Xiaofeng, H., & Ye, J. (2009). A genetic algorithm for minimizing makespan of block erection in shipbuilding. *Journal of Manufacturing Technology Management*, *20*(4), 500–512. Scopus. https://doi.org/10.1108/17410380910953757

Kendall, D. G. (1953). Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics*, *24*(3), 338–354. https://doi.org/10.1214/aoms/1177728975

Kozik, A., & Rudek, R. (2018). An approximate/exact objective based search technique for solving general scheduling problems. *Applied Soft Computing*, *62*, 347–358. https://doi.org/10.1016/j.asoc.2017.10.043

Lawler, E. L., Lenstra, J. K., Rinnooy Kan, A. H. G., & Shmoys, D. B. (1993). Chapter 9 Sequencing and scheduling: Algorithms and complexity. In *Handbooks in Operations Research and Management Science* (Vol. 4, pp. 445–522). Elsevier. https://doi.org/10.1016/S0927-0507(05)80189-6

Li, N., Wu, P., Wang, Y., & Cheng, J. (2022). *Energy-conscious Single-machine Scheduling Problem with Release Dates under Time-of-use Electricity Tariffs*. *2022-December*, 252–256. Scopus. https://doi.org/10.1109/IEEM55944.2022.9989820

Markowitz, H. M. (1991). Portfolio Selection, Efficient Diversification of Investments. *Journal of the Institute of Actuaries*, *119*(1), 165–166. https://doi.org/10.1017/S0020268100019831

Meijer, D. (2022). *Probability Theory for Engineers*. University of Twente.

Meijer, D. (2023). *Statistics for Engineers*. University of Twente.

Ministerie van Volksgezondheid, Welzijn en Sport. (2019, July 15). *Wet verplichte ggz (Wvggz)*. Wet verplichte ggz (Wvggz); Ministerie van Volksgezondheid, Welzijn en Sport. https://www.dwangindezorg.nl/wvggz

Pandey, A., Alexander, D. A., & Kumar, S. K. (2023). *Enhancing Efficiency and Delivery Performance Through Optimization of Machine Scheduling in Pre-emptive Parallel Manufacturing Systems*. 1478–1482. Scopus. https://doi.org/10.1109/IEEM58616.2023.10406554

Paulus, J. J., Ye, D., & Zhang, G. (2009). Optimal online-list batch scheduling. *Information Processing Letters*, *109*(19), 1125–1128. https://doi.org/10.1016/j.ipl.2009.07.006

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*, 240–242.

Pesant, G., Meel, K., & Mohammadalitajrishi, M. (2024). On the Usefulness of Linear Modular Arithmetic in Constraint Programming. In *Integration of Constraint Programming, Artificial Intelligence, and Operations Research* (pp. 248–265). https://www.researchgate.net/publication/352475831_On_the_Usefulness_of_Linear_Modular_Arithmetic_in_Constraint_Programming

Potts, C. N., & Kovalyov, M. Y. (2000). Scheduling with batching: A review. *European Journal of Operational Research*, *120*(2), 228–249. https://doi.org/10.1016/S0377-2217(99)00153-8

Rechtbank Noord-Holland. (2025a, March 19). *Werk- en rechtsgebied rechtbank Noord-Holland*. Rechtspraak.nl. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Noord-Holland/Werk-en-rechtsgebied/Paginas/default.aspx

Rechtbank Noord-Holland. (2025b, May 22). *Rechtbank Noord-Holland*. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Noord-Holland/Paginas/default.aspx

Rocholl, J., & Mönch, L. (2024). Scheduling Jobs in Flexible Flow Shops with s-batching Machines Using Metaheuristics. *International Transactions in Operational Research*, *32*(1), 38–68.

Sheikh, T., Singh, S. K., Kashyap, A., & Shivaji, C. (2013). *APPLICATION OF QUEUING THEORY FOR THE IMPROVEMENT OF BANK SERVICE*. https://www.semanticscholar.org/paper/APPLICATION-OF-QUEUING-THEORY-FOR-THE-IMPROVEMENT-Sheikh-Singh/a02bb55fc3f466fcf31d6195b03deb7c787caa6b

Shen, L., & Buscher, U. (2012). Solving the serial batching problem in job shop manufacturing systems. *European Journal of Operational Research*, *221*(1), 14–26. https://doi.org/10.1016/j.ejor.2012.03.001

Sonenberg, N., Volodina, V., Challenor, P. G., & Smith, J. Q. (2024). Using infinite server queues with partial information for occupancy prediction. *Journal of the Operational Research Society*, *75*(2), 262–277. Scopus. https://doi.org/10.1080/01605682.2023.2189002

Staats, B. R., & Gino, F. (2012). Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science*, *58*(6), 1141–1159. https://doi.org/10.1287/mnsc.1110.1482

Stidham, S. (1974). Technical Note—A Last Word on L = λW. *Operations Research*, *22*(2), 417–421. https://doi.org/10.1287/opre.22.2.417

Su, X., & Zenios, S. (2004). Patient choice in kidney allocation: The role of the queueing discipline. *Manufacturing and Service Operations Management*, *6*(4), 280–301. Scopus. https://doi.org/10.1287/msom.1040.0056

Sundarapandian, V. (2009). *Probability Statistics and Queing Theory 1st Edition V Sundarapandian all chapter instant download | PDF | Stochastic Process | Probability Distribution* (1st ed.). PHI Learning. https://www.scribd.com/document/830373617/Probability-Statistics-and-Queing-Theory-1st-Edition-V-Sundarapandian-all-chapter-instant-download

Tan, T., & Netessine, S. (2014). When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Worker Productivity. *IT & Operations Management Research*. https://scholar.smu.edu/business_itopman_research/2

Teixeira, J. C., Bigotte, J. F., Repolho, H. M., & Antunes, A. P. (2019). Location of courts of justice: The making of the new judiciary map of Portugal. *European Journal of Operational Research*, *272*(2), 608–620.

Whitt, W. (1993). APPROXIMATIONS FOR THE GI/G/m QUEUE. *Production and Operations Management*, *2*(2), 114–161. https://doi.org/10.1111/j.1937-5956.1993.tb00094.x

Winston, W. L. (2022). *Operations Research: Applications and Algorithms*. Cengage Learning.

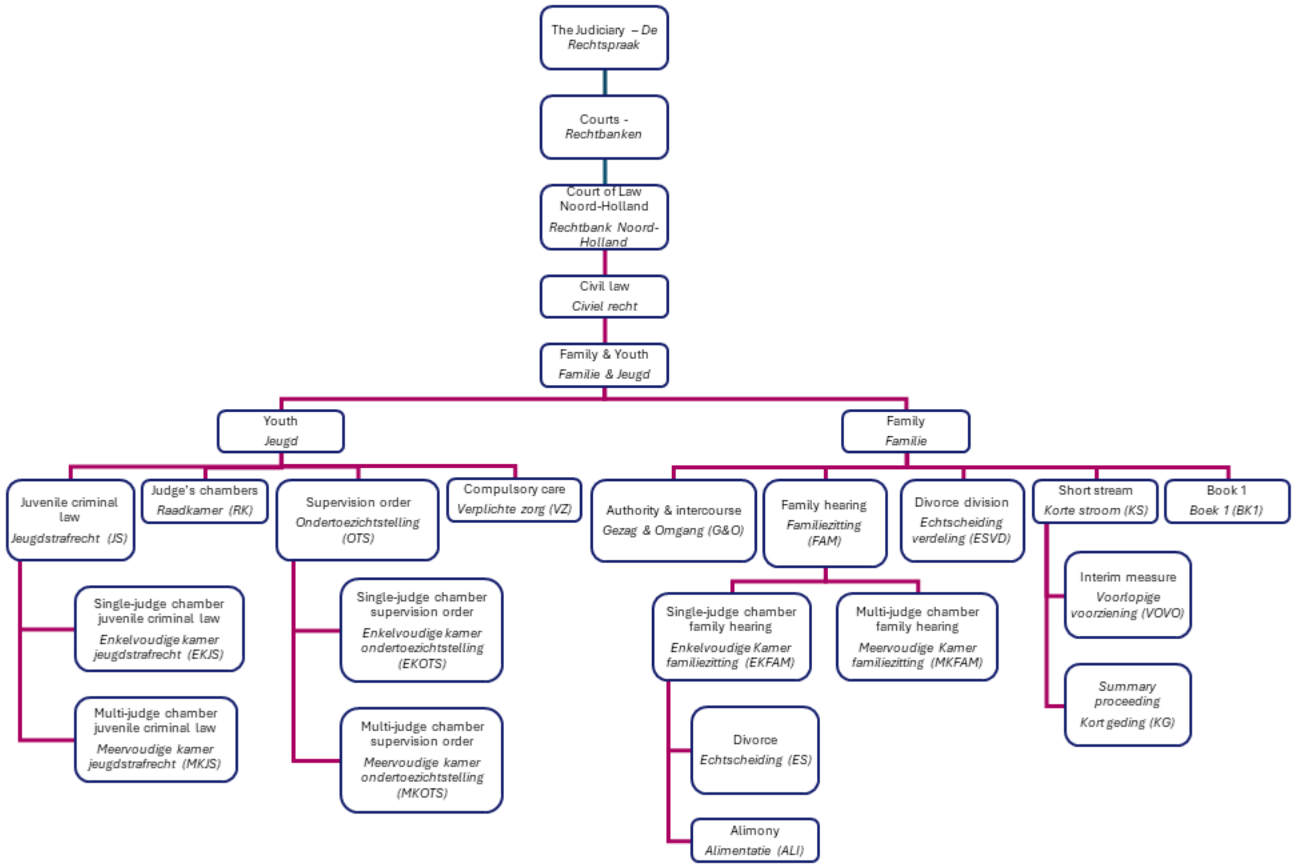
Xie, N., Qin, Y., Chen, N., & Yang, Y. (2025). Single batch-processing machine scheduling problem with interval grey processing time. *Applied Soft Computing*, *170*, 112661. https://doi.org/10.1016/j.asoc.2024.112661

Yang, F., Davari, M., Wei, W., Hermans, B., & Leus, R. (2022). Scheduling a single parallel-batching machine with non-identical job sizes and incompatible job families. *European Journal of Operational Research*, *303*(2), 602–615. https://doi.org/10.1016/j.ejor.2022.03.027

Zou, J., Zhang, Y., & Wang, L. (2012). Minimizing makespan with chain precedence constraints on identical parallel machines. *Advances in Information Sciences and Service Sciences*, *4*(21), 8–14. Scopus. https://doi.org/10.4156/AISS.vol4.issue21.2

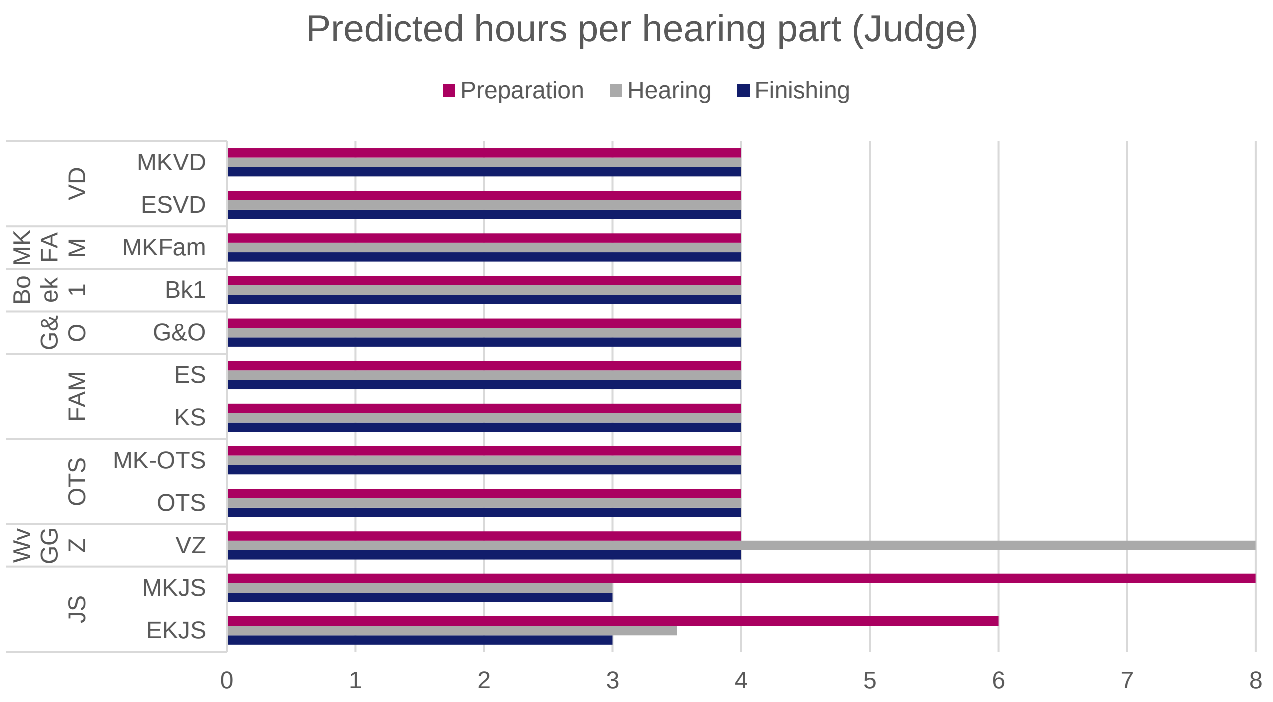
Appendix A Introduction

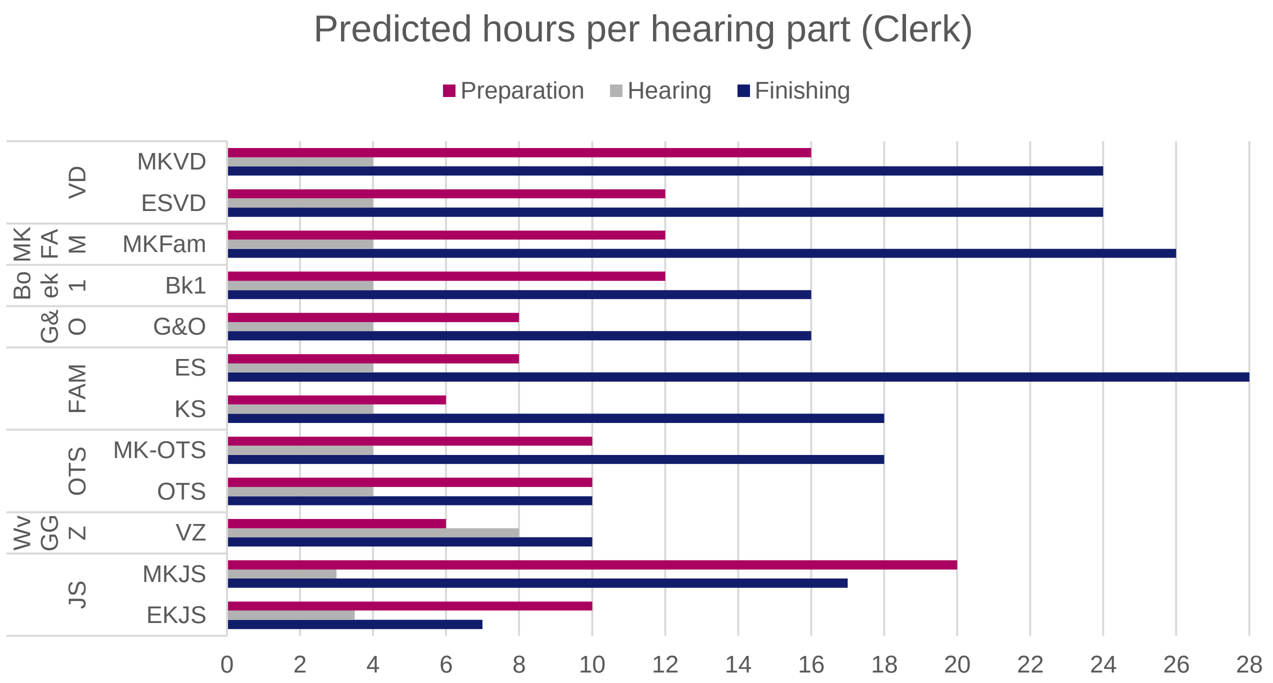
Appendix A.1 F&Y structure

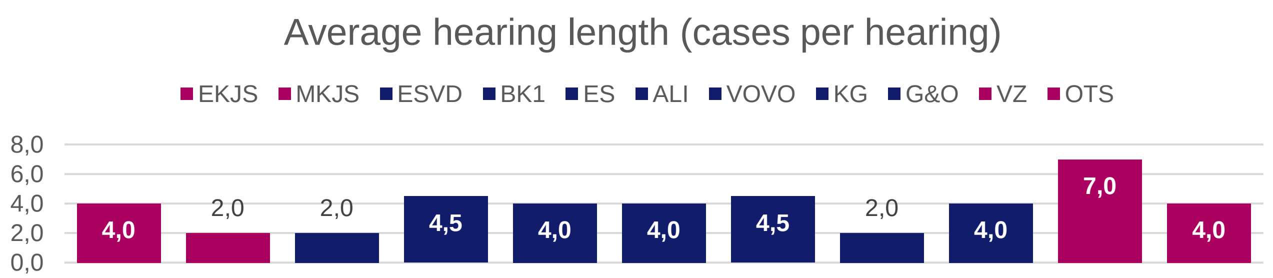


The definition of a *single-judge chamber* (enkelvoudige kamer (EK)) is a hearing where there is one judge supported by one clerk. A *multi-judge chamber* (meervoudige kamer (MK)) is a hearing where there are three judges supported by one clerk. MKs are used for cases where the elaboration of the verdict is difficult due to the complexity of the case.

Appendix A.2 Time prediction

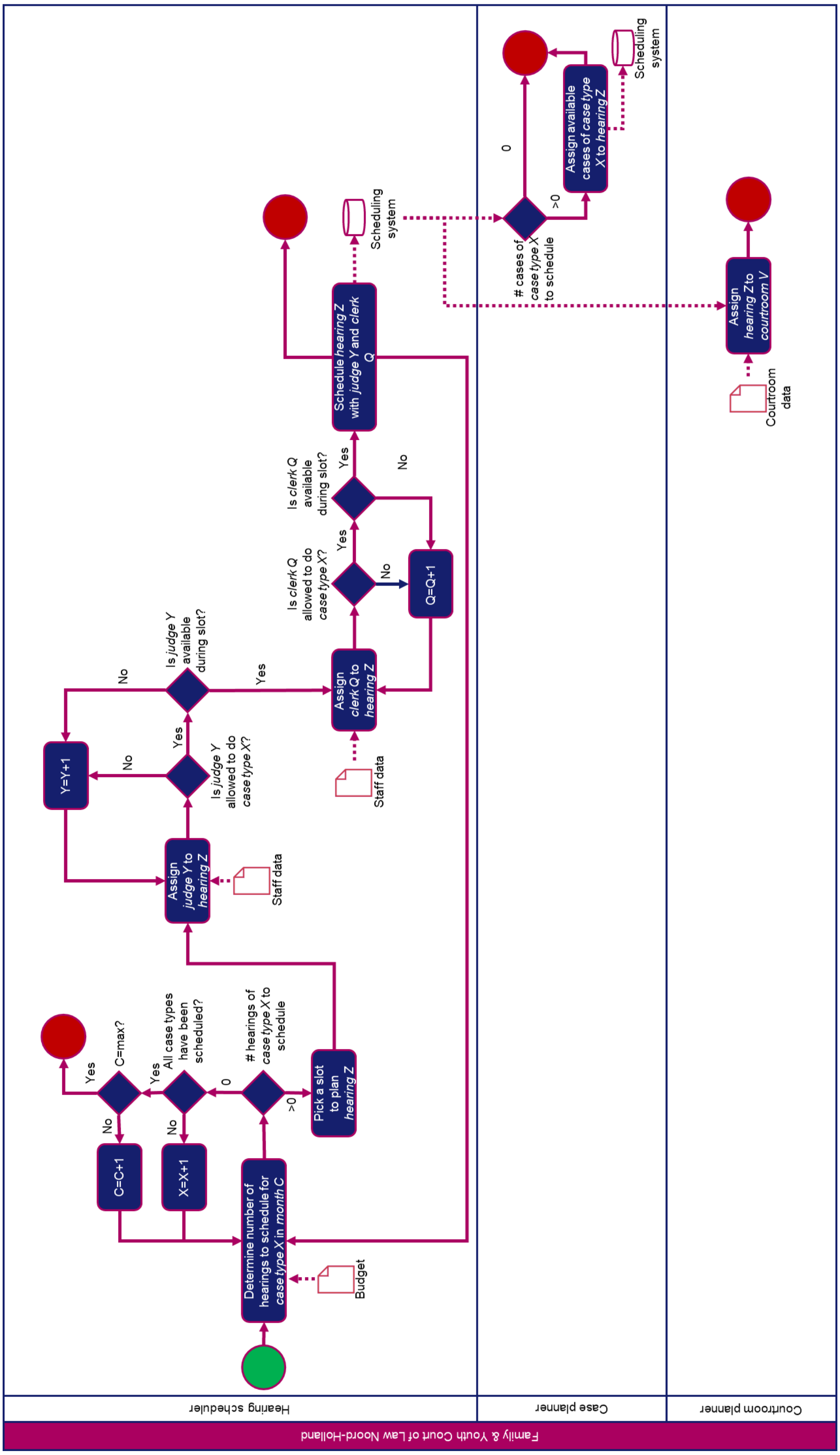




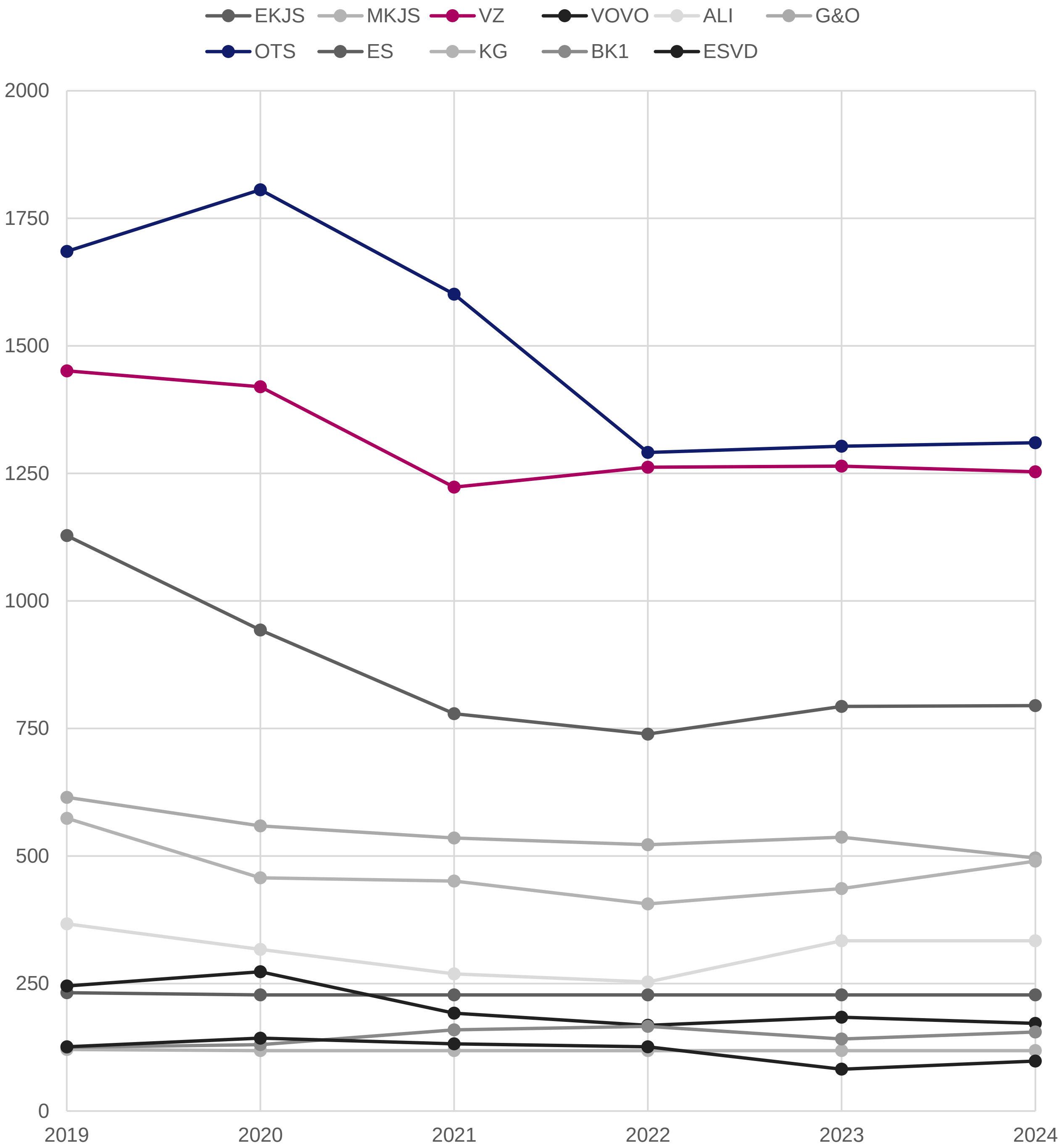


Appendix B Situation analysis

Appendix B.1 BPMN



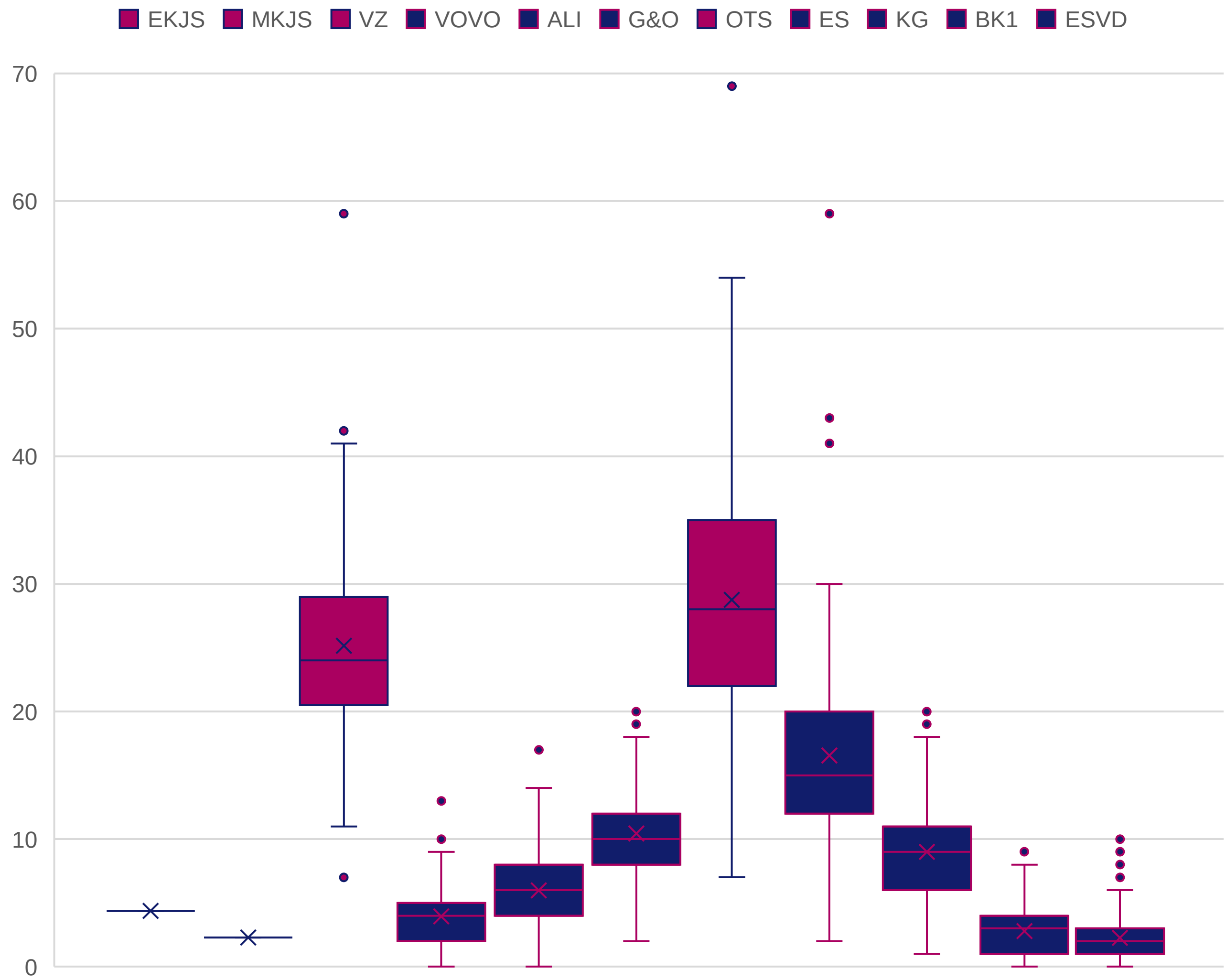
Appendix B.2 Annual case input



We choose to show case input per annum, not per week, for clarity. The OTS case type shows a sudden decrease from 2020 onwards which is caused by changes in law resulting in fewer cases coming on a hearing (de Rechtspraak, 2024a).

The VZ case type input has also been subject to changes in law in 2020 (de Rechtspraak, 2020; Ministerie van Volksgezondheid, Welzijn en Sport, 2019). Trends in other case types do not show a direct causational relationship with changes in law or society that can be assumed with certainty. For example, the ES case type shows a decrease over the years (from around 1,200 to 800 annually). There is no definitive explanation for this. Therefore, it is difficult to predict future case input evolution.

Appendix B.3 Weekly case input per case type



The OTS, VZ, and ES case types are most demanded. Over the years, there have been stark outliers in case input. The Family case types are highlighted in purple; the Youth case types in pink. The case input per case type did not significantly fluctuate year-on-year.

Because there is no weekly input data available for the EKJS and MKJS case types – which are planned by the Public Prosecutors Office (Openbaar Ministerie), a yearly input level of 2024 is used and spread out over the weeks of the years. We assume that there were no changes in case input year-on-year, and that demand entered the system constantly at the same rate. This explains why EKJS and MKJS do not have outlying values.

Below, a table is present with the standard deviations of the case input from 2019 up and until 2024, showing little deviation over the years. Calculation is done with the following formula where x_i is the value of the data points, $\bar{x}$ the mean, and n the number of data points:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

EKJS	MKJS	VZ	VOVO	ALI	G&O	OTS	ES	KG	BK1	ESVD
2,046E-14	8,9E-15	6,709	2,230	2,826	3,463	9,373	6,234	3,564	1,900	8,896E-15

Appendix C Solution development

Appendix C.1 Mathematical model

Sets

$T = the\ set\ of\ all\ primary\ hours\ (t, t' \in \{1,2, \dots |T|\})$

$I = the\ set\ of\ all\ hearing\ type\ specialisations\ \left(i, i' \in \begin{cases}\{1, \dots, 7\}, for\ a\ Family\ judge \\ \{1, \dots, 7\}, for\ a\ Youth\ judge\end{cases}\right)$

$J = the\ set\ of\ all\ hearing\ processes\ that\ can\ be\ scheduled\ (j \in \{1, \dots, \varepsilon_i\})$

Parameters

$p_i = number\ of\ preparation\ hours\ per\ case\ of\ hearing\ type\ i$

$h_i = number\ of\ hearing\ hours\ per\ case\ of\ hearing\ type\ i$

$f_i = number\ of\ finalisation\ hours\ per\ case\ of\ hearing\ type\ i$

$\beta_i = number\ of\ buffer\ hours\ per\ case\ of\ hearing\ type\ i$

$\varepsilon_i = number\ of\ cases\ \ of\ hearing\ type\ i\ to\ schedule\ per\ judge$

$\theta = hearing\ block\ length\ in\ hours\ (\theta = 4)$

$\gamma = integer\ used\ for\ modulo$

$r_t = remainder\ of\ modulo\ operator\ at\ hour\ t\ (r \in \{0, \dots, \theta - 1\})$

$r_t = (t-1) - \theta * \gamma \quad \forall t \in T$

$n_t = remaining\ hours\ at\ hour\ t\ before\ a\ hearing\ block\ break\ (n \in \{1, \dots, \theta\})$

$n_t = \theta - r_t \quad \forall t \in T$

$M^{\theta} = maximum\ hearing\ block\ length$

$M^{x_i} = maximum\ value\ of\ x_i$

Decision variables

$x_i = number\ of\ cases\ to\ schedule\ together\ for\ hearing\ type\ i\ (x \in \{1,2,3,4\})$

$P_{tij} = \begin{cases}1, if\ preparation\ of\ hearing\ process\ j\ of\ hearing\ type\ i\ starts\ at\ hour\ t \\ 0, otherwise\end{cases}$

$H_{tij} = \begin{cases}1, if\ hearing\ of\ hearing\ process\ j\ of\ hearing\ type\ i\ starts\ at\ hour\ t \\ 0, otherwise\end{cases}$

$\delta_{ij} = \begin{cases}1, if\ hearing\ type\ i\ is\ scheduled\ in\ hearing\ process\ j \\ 0, otherwise\end{cases}$

Auxiliary variables

$\omega_{tij} = \ number\ of\ cases\ scheduled\ together\ on\ P_{tij}$

$\psi_{tij} = \ number\ of\ cases\ scheduled\ together\ on\ H_{tij}$

$\phi_{ij} = number\ of\ cases\ of\ type\ i\ scheduled\ in\ j$

Objective function

$Min\ Z = Q$

Constraints

Horizon constraints

$t * P_{tij} + x_i * (f_i + \beta_i + h_i + p_i) \leq |T| \qquad \forall t \in T, \forall i \in I, \forall j \in J$

$t * H_{tij} + x_i * (f_i + \beta_i + h_i) \leq |T| \qquad \forall t \in T, \forall i \in I, \forall j \in J$

Precedence constraints

$$\sum_{t'=1}^{T} \left(t' * H_{t'ij}\right) \geq t * P_{tij} + p_i * \omega_{tij} \quad \forall t \in T, \forall i \in I, \forall j \in J$$

$\omega_{tij} \leq M^{x_i} * P_{tij} \qquad \forall t \in T, \forall i \in I, \forall j \in J$

$\omega_{tij} \leq x_i \qquad \forall t \in T, \forall i \in I, \forall j \in J$

$\omega_{tij} \geq x_i - M^{x_i} * \left(1 - P_{tij}\right) \qquad \forall t \in T, \forall i \in I, \forall j \in J$

$\omega_{tij} \geq 0 \qquad \forall t \in T, \forall i \in I, \forall j \in J$

$$\sum_{t'=1}^{T} \sum_{i'=1}^{I} \left(t' * P_{t'i'j+1}\right) \geq \left(t * H_{tij} + (h_i + f_i + \beta_i) * \psi_{tij}\right) \quad \forall t \in T, \forall i \in I, \forall j \in J \backslash \{|J|\}$$

$\psi_{tij} \leq M^{x_i} * H_{tij} \qquad \forall t \in T, \forall i \in I, \forall j \in J \backslash \{|J|\}$

$\psi_{tij} \leq x_i \qquad \forall t \in T, \forall i \in I, \forall j \in J \backslash \{|J|\}$

$\psi_{tij} \geq x_i - M^{x_i} * \left(1 - H_{tij}\right) \qquad \forall t \in T, \forall i \in I, \forall j \in J \backslash \{|J|\}$

$\psi_{tij} \geq 0 \qquad \forall t \in T, \forall i \in I, \forall j \in J \backslash \{|J|\}$

Hearing block break constraints

$0 \leq r_t < \theta \quad \forall t \in T$

$$n_t - h_i * x_i \geq -M^{\theta} \left(1 - \sum_{i'=1}^{I} \left(H_{ti'j}\right)\right) \quad \forall t \in T, \forall i \in I, \forall j \in J$$

Hearing process planning constraints

$$\sum_{i=1}^{I} \left(\delta_{ij}\right) \leq 1 \quad \forall j \in J$$

$$\sum_{j=1}^{J} \left(\phi_{ij}\right) \geq \varepsilon_i \quad \forall i \in I$$

$\phi_{ij} \leq M^{x_i} * \delta_{ij} \qquad \forall i \in I, \forall j \in J$

$\phi_{ij} \leq x_i \qquad \forall i \in I, \forall j \in J$

$\phi_{ij} \geq x_i - M^{x_i} * \left(1 - \delta_{ij}\right) \qquad \forall i \in I, \forall j \in J$

$\phi_{ij} \geq 0 \qquad \forall i \in I, \forall j \in J$

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (P_{tij} + H_{tij}) \leq 1 \quad \forall t \in T$$

Meeting demand constraints

$$\sum_{t=1}^{T}\left(P_{tij}\right)=\delta_{ij}\quad \forall i\in I,\forall j\in J$$

$$\sum_{t=1}^{T}\left(H_{tij}\right)=\delta_{ij}\quad \forall i\in I,\forall j\in J$$

Objective function constraint

$$\sum_{t=1}^{T}(t+(\,h_i+f_i+\beta_i)*\psi_{tij})\le Q\quad \forall i\in I,\forall j\in J$$

$$\psi_{tij}\le M^{x_i}*H_{tij}\qquad \forall t\in T,\forall i\in I,\forall j\in J$$

$$\psi_{tij}\le x_i\qquad \forall t\in T,\forall i\in I,\forall j\in J$$

$$\psi_{tij}\ge x_i-M^{x_i}*\left(1-H_{tij}\right)\quad \forall t\in T,\forall i\in I,\forall j\in J$$

$$\psi_{tij}\ge 0\qquad \forall t\in T,\forall i\in I,\forall j\in J$$

Optimisation constraints

$$\sum_{t=1}^{T}\sum_{i=1}^{I}\left(P_{ti|J|}\right)=1$$

$$\sum_{t=1}^{T}\sum_{i=1}^{I}\left(H_{ti|J|}\right)=1$$

Appendix C.2 Model code

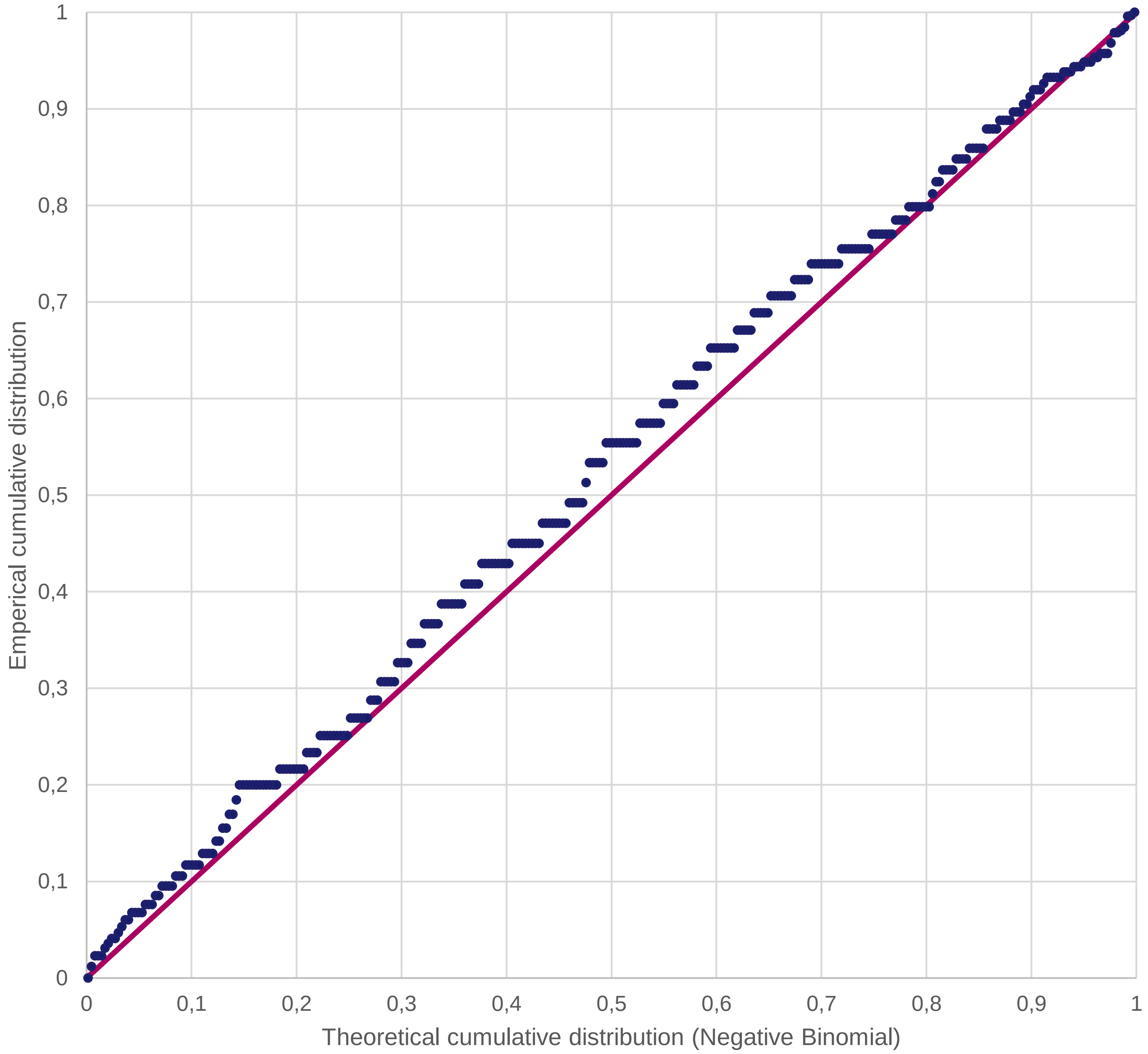
```
import gurobipy as gp
from gurobipy import quicksum, GRB

#Define model
Z = gp.Model()

# -> Parameters <-

epsilon = {(1): 2,
     (2): 1,
     (3): 4,
     (4): 0,
     (5): 0, }  #number of cases to schedule of hearing type i

p = {(1): 1.5,
     (2): 4,
     (3): 1,
     (4): 0,
     (5): 0, }  #number of preparation hours per case of hearing type i

h = {(1): 0.9,
     (2): 1.5,
     (3): 1,
     (4): 0,
     (5): 0, }  #number of hearing hours per case of hearing type i

f = {(1): 0.8,
     (2): 1.5,
     (3): 1,
     (4): 0,
     (5): 0, }  #number of finalisation hours per case of hearing type i

beta = {(1): 0,
     (2): 0,
     (3): 0,
     (4): 0,
     (5): 0, }  #number of buffer hours per case of hearing type i

theta = 4   #Hearing block length (4 hours -> 9:00-13:00 & 13:00-17:00)
M = theta   #Big M

MinimumNeededTime = 0   #Reset minimum needed time
J_range = 0             #Reset the range of J, the set of all available hearing processes

#-> Sets <-
I = range(1,4)    #The set of all hearing type specialisations of a Family/Youth judge
(i,i'∈{1,…,X})

for i in I:
    MinimumNeededTime += epsilon[i] * (p[i] + h[i] + f[i] + beta[i])    #Determine the minimum
time needed to fulfill all cases
    J_range += epsilon[i]                                        #All cases that need to be
scheduled are the maximum number of hearing processes needed

T_range = int(MinimumNeededTime * 1.5)    #Making horizon large enough
```

```
T = range(1,T_range + 1)                  #The set of all primary hours (t,t'∈{1,2,…})
J = range(1,J_range + 1)                  #The set of all hearing processes that can be scheduled
(j∈{1,…,X})

#-> Decision variables <-
lbs = [1,1,1] #Lower bound of hearing length
ubs = [4,2,4] #Upper bound of hearing length
x = Z.addVars(I, lb = lbs, ub = ubs, vtype = GRB.INTEGER, name = "x")   #Main decision variable
(x_i ∈ {1,..,4})
P = Z.addVars(T, I, J, vtype=GRB.BINARY, name = "P")                #1 if preparation of hearing
process j of hearing type i starts at hour t, 0 otherwise
H = Z.addVars(T, I, J, vtype=GRB.BINARY, name = "H")                #1 if hearing of hearing
process j of hearing type i starts at hour t, 0 otherwise

delta = Z.addVars(I, J, vtype=GRB.BINARY, name="delta")             #1 if hearing type i is
scheduled in hearing j, 0 otherwise
makespan = Z.addVar(vtype=GRB.CONTINUOUS, name="makespan")          #Time needed to finish all
cases

omega = Z.addVars(T,I,J, vtype=GRB.INTEGER, name = "omega")
psi = Z.addVars(T,I,J, vtype=GRB.INTEGER, name = "psi")
phi = Z.addVars(I,J, vtype=GRB.INTEGER, name = "phi")
M_x = ubs

#-> Constraints <-

#1. Horizon constraints
for i in I:
    for j in J:
        for t in T:
            Z.addConstr(t * H[t, i, j] + x[i] * (f[i] + beta[i] + h[i]) <= max(T))          #Max
start time of hearing to not surpass horizon
            Z.addConstr(t * P[t, i, j] + x[i] * (f[i] + beta[i] + h[i] + p[i]) <= max(T))   #Max
start time of preparation to not surpass horizon

#2. Precedence constraints
            Z.addConstr((t * P[t, i, j] + p[i] * omega[t,i,j]) <= quicksum(t_prime * H[t_prime, i,
j] for t_prime in T)) #Precedence constraint: hearing j can only start after preparation j has
been completed.
            Z.addConstr(omega[t,i,j] <= M_x[i-1] * P[t,i,j])
            Z.addConstr(omega[t,i,j] <= x[i])
            Z.addConstr(omega[t,i,j] >= x[i]- M_x[i-1] * (1 - P[t,i,j]))
            Z.addConstr(omega[t,i,j] >= 0)

    for j in range(1,J_range):
        for t in T:
           Z.addConstr(quicksum(t_prime * P[t_prime,i_prime,j + 1] for t_prime in T for i_prime in
I) >= (t * H[t,i,j] + (h[i] + f[i] + beta[i]) * psi[t,i,j]))  #Precedence constraint: preparation
j can only start after finalisation j-1 has been completed.)
            Z.addConstr(psi[t,i,j] <= M_x[i-1] * H[t,i,j])
            Z.addConstr(psi[t,i,j] <= x[i])
            Z.addConstr(psi[t,i,j] >= x[i]- M_x[i-1] * (1 - H[t,i,j]))
            Z.addConstr(psi[t,i,j] >= 0)

#3. Hearing block break constraints
for t in T:
```

```
    for i in I:
        for j in J:
            Z.addConstr((theta - ((t - 1) % theta)) - h[i] * x[i] >= -M * (1 - H[t,i,j]))

#4. Hearing process planning constraints
for i in I:
    Z.addConstr(quicksum(phi[i,j] for j in J) >= epsilon[i]) #Make sure that enough hearings
processes are planned per hearing type.
    for j in J:
        Z.addConstr(phi[i,j] <= M_x[i-1] * delta[i,j])
        Z.addConstr(phi[i,j] <= x[i])
        Z.addConstr(phi[i,j] >= x[i]- M_x[i-1] * (1 - delta[i,j]))
        Z.addConstr(phi[i,j] >= 0)

for j in J:
    Z.addConstr(quicksum(delta[i,j] for i in I) <= 1)  #Only one hearing type can be scheduled per
hearing process.

for t in T:
    Z.addConstr(quicksum(P[t,i,j] + H[t,i,j] for i in I for j in J) <= 1)    #No more than one
hearing part can start at time t

#5. Meeting demand constraints
for j in J:
    for i in I:
        Z.addConstr(quicksum(P[t, i, j] for t in T) == delta[i,j])  #\
        Z.addConstr(quicksum(H[t, i, j] for t in T) == delta[i,j])  # > Every part of the hearing
process (preparation, hearing) needs to be scheduled once and only once.


#6. Objective function constraint
for i in I:
    for j in J:
        Z.addConstr(quicksum((t + (h[i] + f[i] + beta[i])) * psi[t, i, j] for t in T) <= makespan)
#Minimises the makespan
        for t in T:
            Z.addConstr(psi[t,i,j] <= M_x[i-1] * H[t,i,j])
            Z.addConstr(psi[t,i,j] <= x[i])
            Z.addConstr(psi[t,i,j] >= x[i]- M_x[i-1] * (1 - H[t,i,j]))
            Z.addConstr(psi[t,i,j] >= 0)

#7. Optimisation constraints
Z.addConstr(quicksum(P[t,i,J_range] for t in T for i in I) ==1 )   #Force J=J_range to be
scheduled
Z.addConstr(quicksum(H[t,i,J_range] for t in T for i in I) == 1)   #Force J=J_range to be
scheduled

#-> Objective function <-
Z.addConstr(makespan >= MinimumNeededTime) #The makespan must be larger than or equal to the
minimum needed time for finishing all hearing process parts
Z.addConstr(makespan <= T_range) #The makespan must be smaller than or equal to the horizon
Z.setObjective(makespan, GRB.MINIMIZE)


Z.setParam("Symmetry", 2)

Z.optimize()
```

```
for t in T:
    for j in J:
        for i in I:
            if P[t, i, j].X > 0.5:
                print(f"P[{t}, {i}, {j}] = {P[t,i,j].x}")
            if H[t, i, j].X > 0.5:
                print(f"H[{t}, {i}, {j}] = {H[t,i,j].x}")
for j in J:
    for i in I:
            if delta[i,j].X > 0.5:
                print(f"delta[{i}, {j}] = {delta[i,j].x}")

for i in I:
    print(f"x[{i}] = {x[i].X}", f"| epsilon[{i}] = {epsilon[i]}")

IdleTime = Z.ObjVal-MinimumNeededTime-1   #Determine the size of the idle time
print('Minimum makespan: %g' % Z.ObjVal, "| Minimum needed time:", MinimumNeededTime, "| Idle
time:", IdleTime)
print("Runtime:",Z.Runtime)
print("T_range:",T_range)
print("J_range:",J_range)

print("IsMIP =",Z.IsMIP)
print("IsQCP =",Z.IsQCP)
print("IsQP =",Z.IsQP)

print("# Quadratic constraints =",Z.NumQConstrs)
print("# Nonzero constraints =",Z.NumNZs)
```

Appendix C.3 P-P plot & PCC



In a *P-P plot*, “two cumulative distribution functions (CDF) [are plotted] against each other” (Gibbons, 2003). With a *P-P plot*, we can assess the closeness of two datasets. On the Y-axis, the CDF of F&Y’s case input is plotted; on the X-axis the CDF of the expected frequency of the *Negative Binomial* distribution.

We use *Pearson’s correlation coefficient (PCC)* to measure the linear correlation between the expected case input frequency under the *Negative Binomial* distribution and the observed case input values of F&Y. The *PCC* value ranges between -1 and 1, indicating a strong negative correlation or a strong positive correlation, respectively. The *PCC* value of F&Y’s case input data set is found to be 0.9927, meaning that there is a very strong positive correlation between the *Negative Binomial* distribution and F&Y’s case input (Pearson, 1895).