UNIVERSITY
OF TWENTE.

MASTER THESIS

“Antecedents and Predictive Modelling of Teacher Absenteeism in Dutch Secondary Education”

By

J.G.H. Engelbertink (s3273172)

Faculty of Behavioural, Management & Social Sciences

MSc. Business Administration

University of Twente

10-04-2025

Examination Committee:

1st Supervisor: dr. M. de Visser

2nd Supervisor: dr.ir. N.J. Pulles

External Supervisor:

X. Hoppenbrouwer (Infotopics B.V.)

ABSTRACT

Purpose – The Dutch secondary education sector experiences significantly higher absence rates compared to the national average. Teacher absenteeism is a crucial challenge for educational organisations, with substantial implications on students' performance, organisational costs, as well as its culture. Given the multi-faceted nature of absenteeism, numerous antecedents influence teacher absenteeism. While teacher absenteeism has been widely studied internationally, research on its antecedents in Dutch secondary education remains lacking, and predictive models tailored to this context are lacking despite growing practical demand. Consequently, it aims to identify key antecedents influencing teacher absenteeism in Dutch secondary education and apply Machine Learning (ML) models to predict absence duration categories (short-term, medium, and long-term). Thereby, providing a more comprehensive understanding of teacher absenteeism.

Method – This study applies a multi-method sequential research design, structured into three interrelated studies. Study 1 consist out of eight qualitative interview with secondary education professionals to identify that are currently not captured by literature. Study 2 statistically analyses antecedents identified from literature and Study 1 using regression analysis (OLS and Poisson) on HR data from 1.150 secondary teachers during the 2023-2024 academic year. Absenteeism was analysed from two perspectives: absence duration and absence frequency, excluding planned maternity or pregnancy leave. These antecedents are categorised into personal, job-related, and socio-economic antecedents. Study 3 applies two ML model (Naïve Bayes and Support Vector Machine), using the significant predictors from Study 2 to predict absence duration categories.

Findings – This research identifies an extensive amount of antecedents influnecing teacher absenteeism in Dutch secondary education, highlighting both established and novel antecedents. Consequently, the 'key' antecedents that consistently influence teacher absenteeism across both perspectives (duration and frequency) seem to be the most important. These include: gender, prior absenteeism, tenure, working hours, team-level absenteeism, and employment status. Additionally, school type and FTE in team were identified as novel antecedents. To predict absence duration categories, this research demonstrates the predictive potential of ML algorithms. The Naïve Bayes model outperformed SVM, and showed that individual absence duration categories can be accurately classified, with an accuracy of 90.7% and F1-score of 0.94.

Contribution – This study provides a more comprehensive understanding of teacher absenteeism by confirming established antecedents and identifying new, context-specific antecedents in Dutch secondary education. Highlighting that absenteeism is a multi-faceted and complex problem. In addition, it demonstrates that ML algorithms, built on validated antecedents, enable individual-level absenteeism predictions. Educational institutions can optimise proactive absenteeism management strategies for high-risk groups to enhance sustainable employability, ultimately reducing absenteeism. This research addresses the need for more comprehensive teacher absenteeism analytics, shifting from descriptive to predictive analytics.

Keywords

Teacher absenteeism, Antecedents, Secondary education, Regression analysis, Machine Learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

High rates of absenteeism pose substantial challenges to organisations (Gosselin et al., 2013; Stewart et al., 2003). In the Netherlands alone, worker absenteeism accounted for €23.3 billion in costs in 2023[1]. Over the past decade, absence rates in the Dutch educational sector have consistently been higher compared to the national average, highlighting the sector's struggle with absenteeism. In the first quarter of 2024, absenteeism rates in education reached 6%, exceeding the average rate of 5.5%[1]. By 'education', the Central Bureau for Statistics (CBS) refers to the entire sector, not only primary and secondary education. The average absenteeism rate in secondary education is even higher, counting 6.2%[2]. Highlighting that teacher absenteeism is a crucial challenge for educational organisations, with significant implications on students' performance, cost of education, and organisational culture (Miller et al., 2008; Scott & Wimbush, 1991; Shapira-Lishchinsky & Rosenblatt, 2010b). Additionally, teacher absenteeism results in teacher shortages, while replacements are difficult to find, possibly less qualified, and less experienced (Leuven, 2008; VfPf, 2024). According to the Dutch Ministry of Education, Culture, and Science, there is a shortage of 8.900 full-time equivalent (FTE) in primary and secondary education[3]. The expectation is that these shortages will remain constant if current conditions remain. Moreover, they found that the permanent deficit subjects are mainly mathematics, physics, chemistry, Dutch, German, French, and Classical languages. These shortages lead to decreases in student productivity when replacement staff is necessary (Green, 2014). Research by Miller et al. (2008) indicates that ten additional days of teacher absence decreases mathematic grades by 3.2% of a standard deviation among fourth-grade students in the United States. These statistics highlight the need to address teacher absenteeism in Dutch secondary education.

The antecedents of absenteeism are extensively researched. Multiple studies found determinants correlating with teacher absenteeism and can be categorised into personal, job-related, and socio-economic antecedents (Harrison & Martocchio, 1998; Knoster, 2016; Miller et al., 2008; Porter & Steers, 1973; Rosenblatt & Shirom, 2005). Despite the extensive research in the United States, Africa, and Asia, literature is lacking on antecedents that

---

[1] CBS. (2024). *Sick leave*. https://www.cbs.nl/nl-nl/visualisaties/dashboard-arbeidsmarkt/werkenden/ziekteverzuim

[2] VfPf. (2024). *Ziekteverzuim in het onderwijs*. https://www.vfpf.nl/ziekteverzuim-in-het-onderwijs

[3] Dutch Ministry of Education, Culture, and Science (2023). *Trendrapportage Arbeidsmarkt Leraren po, vo en mbo 2023*. https://www.aob.nl/assets/Bijlage-3.1-Trendrapportage-Arbeidsmarkt-Leraren-2023.pdf

influence teacher absenteeism in Dutch secondary education. Smulders (1984) published a 30-year absenteeism literature review in which domestic and foreign publications were assessed. The study emphasises that it is challenging to apply research material regarding absenteeism from abroad due to data incompleteness or discrepancies in the operationalisation of absenteeism. While previous literature has focused on a small set of antecedents, they all found some relationship with absenteeism.(Harrison & Martocchio, 1998; Knoster, 2016; Miller et al., 2008; Porter & Steers, 1973; Rosenblatt & Shirom, 2005). Suggesting that absenteeism is a multi-faceted problem. However, there remain unexplored context-specific areas in this field (e.g., school type and employment at multiple schools) However, empirical evidence to support this claim is currently lacking. These gaps highlight a clear need for research on teacher absenteeism in Dutch secondary educational context, using a more comprehensive and context-specific set of antecedents.

In addition to the theoretical gaps, there is also a practical need for improved absenteeism analytics in the educational sector. Infotopics, a data consultancy firm providing HR analytic dashboards to educational institutions, reports increasing client demand for more comprehensive and predictive insights into absenteeism patterns (Hoppenbrouwer & Bouma, 2021). Currently, these educational institutions rely on descriptive analytics, lacking predictive insights. There is a need to develop predictive tools that can inform proactive absence management, particularly given the staffing challenges that the sector faces[4]. These insights are essential to proactively optimise absenteeism management strategies, ultimately reducing absenteeism. Although, previous studies have applied Machine Learning (ML) algorithms to predict absenteeism (Ajmi, 2019; Fernandes & Filho, 2021; Gayathri, 2018; Montano et al., 2020), there remains a lack of research tailored to the educational context.

Consequently, this study aims to identify key antecedents influencing teacher absenteeism in Dutch secondary education and to develop a predictive model for classifying absenteeism categories based on these antecedents. Therefore, this research addresses two key questions: *"What are the domain-specific antecedents of teacher absenteeism in Dutch secondary education?"* and *"How accurately can machine learning classify teacher absence duration categories?"*

[4] Dutch Ministry of Education, Culture, and Science (2023). *Trendrapportage Arbeidsmarkt Leraren po, vo en mbo 2023*. https://www.aob.nl/assets/Bijlage-3.1-Trendrapportage-Arbeidsmarkt-Leraren-2023.pdf

To address these research questions, a multi-method sequential research design is employed. More specifically an exploratory sequential design, to investigate teacher absenteeism in Dutch secondary education that combines qualitative exploration, statistical modelling and predictive Machine Learning (ML) models. Therefore, this research progresses into three distinct studies, presented in Table 1.

Table 1 Multi-Method Sequential Research Design

Study	**Data source**	**Analytical approach**	**Purpose**
Study 1: Qualitative interviews	8 interviews with teachers, principals, and administrators.	GIOIA analysis	Identify underlying antecedents influencing teacher absenteeism, including those not covered in existing literature.
Study 2: Regression analysis	HR data from 1.150 teachers within a Dutch secondary educational institution	OLS regression (log-transformed) for absence duration and Poisson regression for absence frequency	Statistically validate absenteeism antecedents identified in Study 1 and the literature to determine their significance and directionality
Study 3: Machine Learning prediction	HR data from 1.150 teacher within the same Dutch secondary educational institution	Naïve Bayes & Support Vector Machine (SVM) classification	Predict absence duration categories (short-term, medium, long-term) based on significant antecedents from Study 2, enhancing practical applicability.

This multi-method sequential research design combines qualitative and quantitative research methods. Qualitative research is known for its flexibility, adaptability, depth, and discovery (Gioia et al., 2012). While quantitative research is characterised by its focus on collecting and analysing data to examine relationships among variables based on statistics (Creswell, 2014). Study 1 allows for an in-depth exploration of the teacher absenteeism antecedents that are currently not captured by literature (Asamoah, 2014). While Study 2 statistically analyses the relationships between identified antecedents and teacher absenteeism, integrating insights from both the literature review and Study 1. Study 3 adds a practical layer, demonstrating how validated predictors can be used to forecast absence duration categories. The combination of research methods provides a deeper understanding of complex research dynamics and the limitations of a research method are offset by the strengths of another (Östlund et al., 2011). This mixed research design provides a robust, multi-faceted understanding (Creswell, 2014). The dataset in studies 2 and 3, provided by Infotopics via PowerBI, includes Human Resource (HR) analytics from a Dutch secondary educational institution, focused on absence duration and frequency. As these measures are commonly used in absenteeism studies, they are also employed in this study (Gellatly, 1995; Knoster, 2016; Porter & Steers, 1973; Rosenblatt & Shirom, 2005). For both perspectives, absenteeism registration is based on the Dutch NVS (absenteeism statistic) framework, using the 'Wet Poortwachter'[5]. For consistency, this NVS classification is also utilised in Study 3 for absence duration categories: short-term (≤ 7 days), medium (≥ 8 and ≤ 42 days), long-term (≥ 43 and ≤ 365 days), and extra long-term(≥ 366 and ≤ 730 days).

From a theoretical perspective, our findings confirm established antecedents, such as gender, prior absenteeism, working hours, and team-level absence (Harrison & Martocchio, 1998; Knoster, 2016; Miller et al., 2008; Porter & Steers, 1973; Rosenblatt & Shirom, 2005). In addition, it identifies new antecedents of teacher absenteeism (FTE in team and school type). Importantly, the research highlights discrepancies in relation to literature, as well as between absence duration and frequency. Moreover, Study 3 demonstrates that the Naïve Bayes model predict individual absence duration categories of teachers with high accuracy (90.7%), consistent with findings from previous research on absence prediction accuracy in other contexts (Ajmi, 2019; Fernandes & Filho, 2021; Gayathri, 2018; Montano et al., 2020). From a managerial standpoint, educational institutions can optimise proactive absenteeism management strategies for high-risk groups and, thus, enhance sustainable employability, ultimately reducing absenteeism. This research addresses the need to move from descriptive statistics to more comprehensive predictive HR absenteeism analytics. Also, the findings support Infotopics with the commercial ability to offer these HR analytics to educational institutions.

Utilising the classification of antecedents influencing absenteeism from Porter and Steers (1973), the subsequent section of this thesis will begin with an extensive review of existing literature regarding absenteeism, types of absenteeism, its consequences, and antecedents of absenteeism. Subsequently, Study 1 identifies key antecedents of teacher absenteeism through qualitative interviews. Then based on the literature review, findings from Study 1, and available data, hypotheses are developed and tested using regression analysis in Study 2. Afterwards, the identified significant antecedents are used as input variables for Study 3 which employs ML techniques to predict absenteeism categories. The final part discusses the findings across the studies, the implications, limitations, and future research topics.

---

[5] CBS. (2007). *NVS Standaard voor verzuimregistratie:Nationale Verzuimstatistiek*. https://www.cbs.nl

2. LITERATURE REVIEW

This literature review aims to examine and explore current literature on absenteeism, specifically it seeks to understand factors contributing to teacher absenteeism. Understanding teacher absence is essential, research by Miller et al. (2008) argues that teacher absenteeism significantly influences students' performance. Additionally, teacher absenteeism impacts the cost of education and organisational culture (Miller et al., 2008; Scott & Wimbush, 1991; Shapira-Lishchinsky & Rosenblatt, 2010b). Key search terms in academic databases (e.g. Scopus and Web of Sciences), such as 'teacher absenteeism', 'determinants of absenteeism', 'predictive analytics in education', and 'antecedents of absenteeism', were used to examine existing literature.

2.1 DEFINING ABSENTEEISM

Various definitions of absenteeism are reflected across different studies and industries. The concept of absenteeism was first recognised during the industrial era when employers identified absence as an instrument for enhancing efficiency and reducing costs in manufacturing plants (Porter & Steers, 1973). In recent years, studies on absenteeism have evolved by including big data and analytics to identify more in-depth patterns and predictors of absenteeism (Biron & Bamberger, 2010; Knoster, 2016). Schultz et al. (2009) define absenteeism as "the time that an employee is away from work due to illness or disability." Presenteeism, in contrast, refers to employees being present at work but perform below their usual work time. Presenteeism does not necessarily lead to absence (Gosselin et al., 2013; Schultz et al., 2009). Similarly, Johns (2008) describes absence as the inability to attend work as planned, regardless of the reason. While these studies describe absenteeism in general, Rogers and Vegas (2009) define teacher absence as "any instance where a teacher fails to attend school or classroom duties as scheduled."

2.2 TYPES OF ABSENTEEISM

Absenteeism can be differentiated into different types. According to Belita et al. (2013), absenteeism can be classified as anticipated (planned) or without prior notification (unplanned). A second classification is based on absence duration, distinguishing between short-term and long-term. Brief absences related to illness or personal matters are known as short-term absences, whereas long-term absences refer to an extended period away from work, usually due to serious health issues (Schlagman & Kvavilashvili, 2008; Vijayasingham & Mairami, 2018).

Furthermore, absenteeism can be classified into authorised and unauthorised. Schlagman and Kvavilashvili (2008) address that authorised absence is approved by the employers (e.g., planned holidays or medical appointments). On the other hand, unauthorised includes unapproved or unplanned absenteeism. Moreover, they argue that specifying the type of absenteeism is essential for organisations to develop effective absence-reduction strategies. As highlighted earlier by Schultz et al. (2009), an additional type of absenteeism is presenteeism, where employees work below contract hours. Presenteeism can lead to higher costs and reduced productivity compared to absenteeism (Johns, 2010). Teacher absenteeism, as studied in this research, refers to any period during which a teacher is absent from their scheduled duties, excluding planned maternity or pregnancy leave.

Infotopics uses the Dutch NVS (absenteeism statistic) based on ‘Wet Poortwachter’ to determine absence cases and absence trajectories (Hoppenbrouwer & Bouma, 2021). The duration of absence is the number of calendar days from the first day of the absence notification until the recovery data, and therefore does not only concern the days of illness that fall in the reporting period. Whereas frequency, the number of new illness cases per employee within a specific period. Moreover, absenteeism is commonly studied from two perspectives, absence duration and frequency (Gellatly, 1995; Knoster, 2016; Porter & Steers, 1973; Rosenblatt & Shirom, 2005). Additionally, NVS classifies absence duration in four categories: short-term (≤ 7 days), medium leave (≥ 8 and ≤ 42 days), long-term (≥ 43 and ≤ 365 days), and extra long-term (≥ 366 and ≤ 730 days)[6].

All in all, the literature covers various types of absenteeism, including planned/unplanned, short-term/long-term, and authorised/unauthorised. The types of absenteeism overlap with each other. Given that Infotopics provided the data based on the NVS standard, this research continues to use these classification. Moreover, absenteeism will be analysed from two perspectives, absence duration and frequency in Study 2.

---

[6] CBS. (2007). *NVS Standaard voor verzuimregistratie:Nationale Verzuimstatistiek*. https://www.cbs.nl

2.3 CONSEQUENCES OF ABSENTEEISM

Research from Harrison and Martocchio (1998) highlights a 20-year review of absenteeism. They suggest that absenteeism can have different results for different organisations and may not be directly negative. While this study calls for more research, conclusions from other literature suggest that teacher absenteeism poses several consequences. In the Netherlands, teachers usually need to teach classes of absent colleagues (Imants & Zoelen, 1995). Moreover, absenteeism results in teacher shortages, while replacements are difficult to find, possibly less qualified and less experienced (Leuven, 2008). As a result, student productivity decreases when replacement staff is necessary (Green, 2014). This aligns with the findings of Nicholson et al. (2006) who emphasised that student productivity decreases when schools have difficulty finding replacement staff. Furthermore, absenteeism involves costly problems for elementary and secondary education. Costs includes wages for both absent and substitute teachers, extra administrative tasks, and evaluation of substitute teachers (Green, 2014; Miller et al., 2008). As mentioned earlier, Miller et al. (2008) found that ten additional days of teacher absences decreases mathematic grades by 3.2% among fourth-grade students in the United States. Lasty, teacher absenteeism is connected to student absenteeism. Educational institutions with higher rates of student absences tend to have increased rates of teacher absences (Ehrenberg et al., 1991).

2.4 ANTECEDENTS OF ABSENTEEISM

In the context of HR, several antecedents influence absenteeism in organisations. Imants and Zoelen (1995) found that less than 20% of absences among Dutch teachers account for sickness absence, suggesting that teacher absence goes beyond sickness. Research from Porter and Steers (1973) stated that employees are absent due to personal, job-related, and socio-economic factors. For the purpose of clarification, the framework proposed by Porter and Steers (1973) to classify antecedents influencing absenteeism is used in this literature review.

2.4.1 Personal antecedents

Various studies have addressed the relationship between personal antecedents and absenteeism. In earlier studies, absenteeism was primarily focused on personal characteristics, such as gender, age, and marital status (Harrison & Martocchio, 1998; Scott & Wimbush, 1991; Steers & Rhodes, 1978). In recent years, studies addressed prior absenteeism, education, children, and ethnicity in relation to absenteeism (Green, 2014;

Knoster, 2016; Nedungadi et al., 2017). The personal antecedents include *age, gender, marital status, education, and prior absenteeism*.

Age

Studies regarding the age of employees influencing absence found different results. Hackett (1990) addressed that older employees tend to be less productive than younger employees. Additionally, the study found a negative correlation between age and avoidable absenteeism. Nevertheless, a positive association between age and absence frequency was found. Similarly, Garcia (1987) found a positive relationship between age and absenteeism. However, the studies by Ann-Kristina and Nielsen (2008) and Gerstenfeld (1969) also addressed a negative relationship. While these studies are conducted in different research contexts, they provide insights into the relationship between age and teacher absenteeism. In the context of education, a negative relationship is mostly found (Harrison & Martocchio, 1998; Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991).

Gender

Research on the relationship between gender and absenteeism found similar results. A longitudinal study by Keller (1983) aimed to predict absenteeism from prior absenteeism, job attitudes, demographic variables, and personality variables. The study found that gender is positively related to absenteeism. Garcia (1987) also found a positive relationship between gender (women) and absenteeism. Although this study was conducted among nurses, it provided proficient insights into the relationship. In education, most teachers tend to be women and are generally more absent than male teachers (Podgursky, 2003). Scott and Wimbush (1991) found that female teachers in secondary education in the United States are more frequently absent than men. Also, women were absent more days than men. Key causes for this divergence are pregnancy, menstruation-related problems, and home responsibilities, such as childcare and supporting other family needs (Drago & Wooden, 1992). In contrast, Martocchio (1989) identified a negative relationship for male teachers but not for women. Nedungadi et al. (2017) also identified gender as a significant antecedent of teacher absenteeism in India. Suggesting that primary reasons for divergence are responsibilities at home, such as taking care of children and supporting other family needs.

Marital status

While some studies found a significant relationship between marital status and absenteeism, others have not found associations (Allebeck & Mastekaasa, 2004; Borman & Dowling, 2008; Keller, 1983; VandenHeuvel & Wooden, 1995). Keller (1983) found a negative relationship between marital status (married) and absenteeism. As mentioned earlier, most teachers are women and are generally more absent than male teachers (Podgursky, 2003). This aligns with the findings of Borman and Dowling (2008), who suggest that young, female, married, and white teachers are more likely to leave teaching in comparison to older, male, unmarried, and non-white (Green, 2014). Additionally, Zuba and Schneider (2013) found a positive relationship between being married and absenteeism. Whereas, VandenHeuvel and Wooden (1995) did not find that marital status altered any significant changes in absenteeism. Moreover, Allebeck and Mastekaasa (2004) also found no empirical evidence between marital status and absence.

Education

Several studies found a significant relationship between education and absence rates. Harrison and Martocchio (1998) addressed in a 20-year literature review of absenteeism that the level of education was commonly found to be a significant predictor. Moreover, Rosenblatt and Shirom (2005) also found that education is a significant predictor of absenteeism frequency, accounting for 50 percent of its variance. Research by Nedungadi et al. (2017) pointed out that parents' educational level can also influence student absenteeism. In the Netherlands, teachers should have at least a pedagogical academy for primary education (PABO) diploma to teach primary education and a first- or second-degree qualification for secondary education[7].

Prior absenteeism

Keller (1983) found a positive relationship between absenteeism and prior absenteeism. A study by Cohen and Golan (2007) also found a strong effect on prior absenteeism. While the study by Keller (1983) was conducted in a manufacturing plant, and the study by Cohen and Golan (2007) among nurses in Israel, it suggests that prior absenteeism influences

[7] Rijksoverheid. (2024). What educational qualifications are there? https://www.rijksoverheid.nl/onderwerpen/werken-in-het-onderwijs/vraag-en-antwoord/welke-onderwijsbevoegdheden-zijn er#:~:text=Het%20diploma%20kunt%20u%20halen,en%201%20voor%20de%20pabo.

absenteeism. In the context of teachers, Rosenblatt and Shirom (2005) also found prior absenteeism to be a significant predictor of teachers in the Israeli public education system.

Children

Research indicates that having children positively influences absence rates (Ehrenberg et al., 1991; Gerstenfeld, 1969; Wonders, 2021; Zuba & Schneider, 2013). Research by Vistnes (1997) revealed a positive relationship between the number of children under the age of six and the likelihood of absence among women. Moreover, Zuba and Schneider (2013) found a positive relationship between having children and absenteeism. These findings align with research by Ehrenberg et al. (1991), which found a positive relationship between having more children and teacher absenteeism rates. On the other hand, Cohen and Golan (2007) found no relationship between having children under eighteen and a higher absence rate. Furthermore, women missed more workdays to care for sick family members or when they have dependent children, indicating that women are more likely to act as caregivers than men (Edwards, 2014; Kamphuis, 2018).

Health status

Health status is referred as the physical and mental well-being of an employee (Kamphuis, 2018). Research by Melander et al. (2022) addressed that specifically physical and mental health problems are antecedents for teacher absenteeism. Multiple studies have found positive relationships between health status and absenteeism. Research by Goldberg and Waldman (2000) found a positive relationship between health and absenteeism among hospital employees in the United States. This aligns with the findings of Dwomoh and Moses (2020), who also found a significant positive correlation among hotel-staff. Additionally, it was found that a high Body Mass Index (BMI) and smoking behaviour increase absence rates (Vuorio et al., 2019; Wonders, 2021). This is consistent with the findings of Harrison and Martocchio (1998), who found that health-related factors such as drug use, smoking, and drinking have strong associations with absenteeism. Conversely, Cohen and Golan (2007) found no evidence to support this relationship. The results above suggest that poor health contributes to higher absenteeism rates.

Menopause

Menopause can influence the health-status among women, ultimately contributing to absenteeism. Research indicates that menopausal symptoms among women have less concentration, tiredness, poor memory, and lower confidence (Geukes et al., 2016; O'Neill et al., 2023; Verdonk et al., 2022). A cross-sectional study among 407 Irish hospital workers found that 18% had taken sick leave due to menopausal symptoms. To address these issues it was highlighted that manager awareness and flexible working times can mitigate absenteeism (O'Neill et al., 2023). This is in line with the findings from Geukes et al. (2016), who found that symptomatic Dutch women reported lower work ability than their counterparts. Leading to more prolonged sickness absence from work.

2.4.2 Job-related antecedents

Various studies have addressed the relationship between job-related antecedents and absenteeism. These antecedents include *tenure, job satisfaction, working environment/conditions, workload, job involvement, commuting distance, stress, leadership, and firm size*.

Tenure

Several studies have highlighted the positive relationship between tenure and absenteeism (Harrison & Martocchio, 1998; Nedungadi et al., 2017; Scott & Wimbush, 1991). Moreover, Garrison and Muchinsky (1977) found that tenure is negatively correlated with unpaid absenteeism, whereas it is positively correlated with paid absenteeism in an accounting department in the US. According to the Dutch Ministry of Education, Culture, and Science, teachers in the Netherlands have 12 months of paid leave[8]. In the context of education, Scott and Wimbush (1991) mentioned that tenure significantly influences teacher absenteeism. This aligns with research by Harrison and Martocchio (1998), who also found a positive relationship.

Experience

Asiyai (2017) found a significant relationship between experiences (years working as a teacher) and teacher absenteeism. The finding of this study indicates that more experienced teachers were more absent than less experienced teachers. These findings align with the study by Ost and Schiman (2017), who found that teachers are less likely to be absent when they have fewer years of experience. Similarly, Speas (2010) found that teachers with less than

[5] Dutch Ministry of Education, Culture, and Science (2023). *Trendrapportage Arbeidsmarkt Leraren po, vo en mbo 2023*. https://www.aob.nl/assets/Bijlage-3.1-Trendrapportage-Arbeidsmarkt-Leraren-2023.pdf

three years of experience had the lowest absenteeism (7.96 days), while those with 4-9 years had the highest (10.65 days). Lastly, teachers in the United States with over ten years of experience were, on average, absent four days more than first-year teachers (Saenz-Armstrong, 2020).

Job satisfaction

Multiple studies found significant relationships between job satisfaction and absenteeism (Green, 2014; Harrison & Martocchio, 1998; Jacob, 2010; Scott & Wimbush, 1991; Steers & Rhodes, 1978). Low job satisfaction leads to higher absenteeism rates (Jaarsveld & Keyser, 2018). This aligns with the findings of Green (2014), who demonstrated that stress and job satisfaction influenced teacher absenteeism. Additionally, Cohen and Golan (2007) found that job satisfaction negatively influenced absenteeism, indicating that lower job satisfaction increases absenteeism. The study by Hendriks (2024), involving data from over 15.000 employees in the Netherlands, found a significant relationship between job satisfaction and absenteeism. The findings indicate that a 0.25-point increase in employees' self-reported job satisfaction reduces one working day in absenteeism per FTE annually. It addresses that employee satisfaction is driven by 48 distinct aspects, with safety and security and work-life balance as the most important determinants.

Working environment/conditions

Building on the work of current literature regarding teacher absenteeism, research by Knoster (2016) further explores determinants of absence. It argues that working conditions, climate, community conditions, and cultural responsibilities appear to influence teacher absenteeism. Similarly, Shapira-Lishchinsky and Rosenblatt (2010b) emphasise the importance of the ethical climate in schools influencing teacher's absence. The study found that teachers who feel attached to their school are more likely to minimise their absence. Moreover, Rosenblatt and Shirom (2005) also highlighted a negative relationship between the working environment and teacher absenteeism. This is consistent with the findings of Bridges and Hallinan (1978), who found that organisational climate is strongly correlated to teacher absenteeism. Additionally, schools with high developed infrastructure facilities reduce the probability of teacher absences by 5 to 7.5 percentage points (Kremer et al., 2005).

Working hours

Studies regarding the working hours of employees influencing absence found different results. While the study by Ost and Schiman (2017) found that absence rates among teachers decrease when they have larger classes, increased workload, and new classes (compared to the previous year). In contrast, Rosenblatt and Shirom (2005) discovered that teachers who are frequently absent tend to work longer work hours, indicating a positive relationship. Moreover, Iverson et al. (1998) observed a positive relationship between working hours and absenteeism among public hospital employees. Additionally, a study by Jacob (2010) found that higher teacher absences were observed on Mondays and Fridays. While this study predominantly focused on African-American high schools, it suggests that day of absence can influence absenteeism. Furthermore, working overtime negatively impacts health-status, thereby increasing absenteeism levels (Tadesse et al., 2015; Virtanen et al., 2011). Specifically, Virtanen et al. (2011) found that women moderate working hours, indicating a stronger effect for women compared to men.

Job involvement

According to Kanungo (1982), job involvement refers to “a cognitive state of psychological identification with the job”. Several studies have argued the role of job involvement in influencing absenteeism. Farrell and Stamm (1988) argued that job involvement for females are significant predictors of absence frequency, indicating that low job involvement results in higher absence rates. Moreover, in the 20-year literature review of absenteeism, Harrison and Martocchio (1998) identified that job involvement and absenteeism have well-documented significant relationships. Additionally, research by Scott and Wimbush (1991) indicates that job involvement negatively relates to absenteeism among teachers in secondary education. At the same time, Cheloha and Farr (1980) found that absenteeism and job satisfaction are mediated by job involvement.

Autonomy

Teacher autonomy refers to the professional independence to take own decision in terms of teaching, which enables to perform unbiased actions for creating learning experiences (Bhushan, 2018). Studies have found (direct and indirect) negative relationship between autonomy and absenteeism. First of all, Collie (2021) found that Australian teachers who experienced autonomy, were less likely to be absent. This is consistent with the findings of Bridges and Hallinan (1978), who highlighted that in schools with high absenteeism rates, teachers experienced less autonomy and flexibility. Similarly, Peng and Guo (2022) identified a significant positive link between teacher autonomy and mental health. The underlying mechanism for this link are that autonomy enhances motivation, job satisfaction, and well-being, ultimately contributing to lower absenteeism levels. Based on the results above, it is suggested that when teachers perceive higher levels of absenteeism, they are less likely to be absent.

Commuting distance

A study by Ommeren and Gutiérrez-i-Puigarnau (2011) revealed that commuting distance strongly affects absenteeism among German workers. They argue that if all employees in the economy have negligible commutes, absenteeism would be approximately 15 to 20% lower. In contrast, VandenHeuvel and Wooden (1995) discovered that commuting time exhibits no significance with absence, although once aggregated by sex, a significant correlation is found for women. Nevertheless, Scott and Wimbush (1991) also found that teacher absenteeism was positively related to distance to work.

Stress

According to Lazarus and Folkman (1984), stress arises when “individuals appraise a situation as threatening or challenging, and they perceive their coping resources as insufficient to manage the demands.” Research has consistently highlighted the relationship between stress and teacher absenteeism. Miller et al. (2008) discovered that teachers take more discretionary days when they experience stressful situations. This is supported by research by Green (2014) and Gaziel (2004), who found a positive relationship between teacher absenteeism and job-related stress. Moreover, Ismail (2023) addressed that emotional stress among teachers in Malaysia is significantly affected by lack of resources, time constraints, and negative student behaviour.

Leadership style

Studies have highlighted that leadership styles influence absence rates (Hassan et al., 2014; Jung & Takeuchi, 2010; Price, 2012). Research by Hassan et al. (2014) highlighted that absenteeism rates are lower when employees perceive leadership as ethical. Leaders encouraging others to act similarly to co-workers will, in turn, increase job satisfaction (Hassan et al., 2014). Moreover, Jung and Takeuchi (2010), found that supportive leadership negatively influences absenteeism. While this study focused on manufacturing organisations in Japan, these findings align with the research by Bowers and White (2014). They found that effective school leadership is associated with lower rates of teacher absenteeism. Moreover, Imants and Zoelen (1995) discovered that teacher-perceived restrictive behaviour by school principals led to higher absence rates. Studies suggest that positive school leadership can play a critical role in the organisational environment, which, in turn, influences teachers' presence (Green, 2014; Owen, 2010). In contrast, the study by Cunningham (2019) found no significant relationships between school principals exhibiting leadership style and teacher absenteeism. Lastly, another study found that transformational leadership in one year led to positive absenteeism in the following year (Nielsen & Daniels, 2016).

Additional responsibilities

Research on the relationship between management positions or additional responsibilities and teacher absenteeism shows discrepancies (Carrizosa & Witte, 2024). Some studies have found that head teachers are more likely to be absent compared to regular teachers (Kremer et al., 2005; Usman et al., 2007). In contrast, other studies suggest that higher-rank positions are associated with lower absenteeism due to more responsibilities and increased motivation to attend school (Banerjee et al., 2012; Bradley et al., 2007). As a result, while there is evidence of a significant relationship, the direction remains uncertain.

Monitoring & Inspection

Several studies found significant relationships between monitoring & inspection and teacher absenteeism. Specifically, a study by Cilliers et al. (2016) highlighted that more frequent monitoring of teacher attendance is associated with lower teacher absence among Ugandan primary schools. Similarly, other studies have emphasised that supervision and monitoring initiatives can reduce teacher absenteeism by approximately 6 to 8 percentage points (Carrizosa & Witte, 2024; Muralidharan et al., 2017; Rogers et al., 2004).

Managerial absence

Research on the relationship between managerial absence shows discrepancies (Ann-Kristina & Nielsen, 2008; Duff et al., 2015; Rentsch & Steel, 2003). Ann-Kristina and Nielsen (2008) found a significant positive correlation between employees' absence frequency and the absenteeism frequency of their managers. Whereas, Duff et al. (2015) found no evidence to suggest that manager absence predicts individual absenteeism. Although these studies are conducted in different settings, it provides valuable insights into the potential relationship.

Team-level absence

Research suggest that team absence significantly influences individual absenteeism, where individuals replicate the behaviour of their team. Moreover, the study found that manager absence moderates this relationship, only when manager absence is low (Duff et al., 2015). This aligns with the findings from Steers and Rhodes (1978) who highlight that employees develop resentment when a tolerant group absence norm is present. If no barriers exist, employees tend to adjust their personal absence norm with the group's absence norm (Geurts et al., 1994).

2.4.3 Socio-Economic antecedents

Literature has highlighted the role of socio-economic antecedents that are associated with absenteeism. The socio-economic antecedents contain *salary* and *contract type*.

Salary

Research on the relationship between salary and teacher absenteeism found similar results. The studies by Pfeifer (2010) and Brooke and Price (1989) both highlighted a negative relationship between pay structure and absenteeism. While both studies focus on different contexts, the findings align with the study by Scott and Wimbush (1991). They also discovered a negative relationship between pay and teacher absenteeism, indicating that higher-paid teachers show lower absence rates. Similarly, Knoster (2016) found that pay structure influences teacher absenteeism in the United States. To supplement their lower salary, they seek for additional income sources, leading to absenteeism (Jean-Francois, 2023).

Contract type

The relationship between contract type and absenteeism has been explored in several studies. Research by Baydoun et al. (2015) argues that the antecedent contract type influences absenteeism rates. While this study was focused on nurses, it aligns with the study by Arai and Thoursie (2005) on private sector organisations in Sweden, which revealed a negative relationship between temporary contracts and absenteeism among private sector organisations in Sweden. Specifically focusing on teachers, Senou (2021) discovered that contract teachers are more frequently absent than those who are on permanent contracts. Moreover, Bradley et al. (2007) found that temporary teachers in Australia take 22-24% fewer days of absence compared to permanent workers. Suggesting that contract teachers benefit from greater job security and experience less pressure regarding attendance at work.

Employment status

Studies on the relationship between employment status (full-time vs. part-time) and absenteeism found similar results. First of all, Brown (1999) found, in a longitudinal study among employees of a manufacturing company in the United Kingdom, that full-time contracts are negatively related with absenteeism. This aligns with the findings from Zeytinoglu et al. (2004), who found that part-time employees are significantly more absent compared to full-time employees. While these studies are conducted in different contexts, it suggest absenteeism differences.

In conclusion, the discussed literature highlights extensive research conducted in the field of absenteeism, both in general and specifically among teachers. Studies have emphasised the impact of teacher absenteeism on student performance, educational costs, and organisational culture (Miller et al., 2008; Scott & Wimbush, 1991; Shapira-Lishchinsky & Rosenblatt, 2010b). Despite the extensive research in the United States, Africa, and Asia, literature is lacking on teacher absenteeism in Dutch secondary education. While previous literature has highlighted personal, job-related, and socio-economic factors influencing absenteeism (Green, 2014; Harrison & Martocchio, 1998; Imants & Zoelen, 1995; Miller et al., 2008; Porter & Steers, 1973; Rosenblatt & Shirom, 2005), there also remain potentially unexplored teacher absenteeism antecedents. Based on the literature, the antecedents that may influence teacher absenteeism in Dutch secondary education have been summarised in Table 2. As earlier mentioned, this research uses the framework from Porter and Steers (1973) to classify categories of absenteeism.

Table 2 Antecedents of Absenteeism

	Antecedent	Study	Result
Personal	**Age**	(Hackett, 1990)	+/-
		(Garcia, 1987)	+
		(Ann-Kristina & Nielsen, 2008)	-
		(Gerstenfeld, 1969)	-
		(Rosenblatt & Shirom, 2005)	-
		(Scott & Wimbush, 1991)	-
	Gender	(Keller, 1983)	+
		(Garcia, 1987)	+ *(female)*
		(Podgursky, 2003)	x
		(Scott & Wimbush, 1991)	+ *(female)*
		(Martocchio, 1989)	- *(male)*
		(Nedungadi et al., 2017)	x
	Marital status	(Keller, 1983)	- *(married)*
		(Borman & Dowling, 2008)	+ *(married)*
		(Zuba & Schneider, 2013)	+ *(married)*

Table 2 Cont.

	Antecedent	Study	Result
Personal	**Marital status**	(VandenHeuvel & Wooden, 1995)	x
		(Allebeck & Mastekaasa, 2004)	x
	Education	(Harrison & Martocchio, 1998)	+
		(Rosenblatt & Shirom, 2005)	x
		(Nedungadi et al., 2017)	x
	Prior absenteeism	(Keller, 1983)	+
		(Cohen & Golan, 2007)	+
		(Rosenblatt & Shirom, 2005)	+
	Children	(Ehrenberg et al., 1991)	+
		(Zuba & Schneider, 2013)	+
		(Gerstenfeld, 1969)	+
		(Cohen & Golan, 2007)	x
	Health status	(Goldberg & Waldman, 2000)	-
		(Dwomoh & Moses, 2020)	-
		(Vuorio et al., 2019)	-
		(Harrison & Martocchio, 1998)	-
		(Cohen & Golan, 2007)	x
	Menopause	(Geukes et al., 2016)	x
		(Verdonk et al., 2022)	-
		(O'Neill et al., 2023)	-
Job-related	**Tenure**	(Harrison & Martocchio, 1998)	+
		(Nedungadi et al., 2017)	+
		(Scott & Wimbush, 1991)	+
		(Garrison & Muchinsky, 1977)	- *(unpaid)*
		(Garrison & Muchinsky, 1977)	+ *(paid)*
	Experience	(Asiyai, 2017)	+
		(Ost & Schiman, 2017)	+
		(Speas, 2010)	+
		(Saenz-Armstrong, 2020)	+
	Job satisfaction	(Green, 2014)	-
		(Harrison & Martocchio, 1998)	-
		(Scott & Wimbush, 1991)	-
		(Porter & Steers, 1973)	x
		(Cohen & Golan, 2007)	-
	Working environment/ conditions	(Knoster, 2016)	-
		(Rosenblatt & Shirom, 2005)	-
		(Shapira-Lishchinsky & Rosenblatt, 2010b)	-
		(Bridges & Hallinan, 1978)	-
		(Kremer et al., 2005)	-
	Working hours	(Ost & Schiman, 2017)	-
		(Rosenblatt & Shirom, 2005)	+
		(Iverson et al., 1998)	+
		(Virtanen et al., 2011)	+ (women)
		(Tadesse et al., 2015)	+
	Job involvement	(Farrell & Stamm, 1988)	- *(female)*
		(Harrison & Martocchio, 1998)	-
		(Scott & Wimbush, 1991)	-
		(Cheloha & Farr, 1980)	x *(mediating)*
	Autonomy	(Collie, 2021)	-
		(Bridges & Hallinan, 1978)	-
		(Peng & Guo, 2022)	-
	Commuting distance	(Ommeren & Gutiérrez-i-Puigarnau, 2011)	+
		(VandenHeuvel & Wooden, 1995)	+ *(female)*
		(Scott & Wimbush, 1991)	+

Table 2 Cont.

	Antecedent	Study	Result
Job-related	**Stress**	(Miller et al., 2008)	+
		(Green, 2014)	+
		(Gaziel, 2004)	+
	Leadership style	(Hassan et al., 2014)	- *(ethical)*
		(Jung & Takeuchi, 2010)	- *(supportive)*
		(Bowers & White, 2014)	-
		(Imants & Zoelen, 1995)	+ *(restrictive)*
		(Green, 2014)	- *(positive)*
		(Bowen, 2009)	- *(positive)*
		(Cunningham, 2019)	x
		(Nielsen & Daniels, 2016)	- *(transformational)*
	Additional responsibilities	(Kremer et al., 2005)	+
		(Usman et al., 2007)	+
		(Banerjee et al., 2012)	-
		(Bradley et al., 2007)	-
	Monitoring/Inspection	(Cilliers et al., 2016)	-
		(Muralidharan et al., 2017)	-
		(Rogers et al., 2004)	-
	Managerial absence	(Ann-Kristina & Nielsen, 2008)	+
		(Duff et al., 2015)	x
	Team-level absence	(Steers & Rhodes, 1978)	+
		(Duff et al., 2015)	+
Socio-economic	**Salary**	(Pfeifer, 2010)	-
		(Brooke & Price, 1989)	-
		(Scott & Wimbush, 1991)	-
		(Knoster, 2016)	-
	Contract type	(Baydoun et al., 2015)	x
		(Arai & Thoursie, 2005)	- *(temporary)*
		(Senou, 2021)	- *(temporary)*
		(Bradley et al., 2007)	- *(temporary)*
	Employment status	(Zeytinoglu et al., 2004)	+ *(part-time)*
		(Brown, 1999)	- *(full-time)*

Note: '+' indicates a positive relationship, '-' a negative relationship, and 'x' means no relationship was found

3. STUDY 1 - QUALITATIVE INTERVIEWS

Figure 1 illustrates the overall multi-study approach adopted in this research. The goal of Study 1 (highlighted in blue) is to identify, through qualitative interviews, potential antecedents that may be context-specific or previously overlooked in academic research. The findings from this study also informs Study 2, where the identified antecedents are quantitatively tested through regression analysis. Subsequently, the significant antecedents are incorporated into Study 3, where ML algorithms predict absenteeism categories.

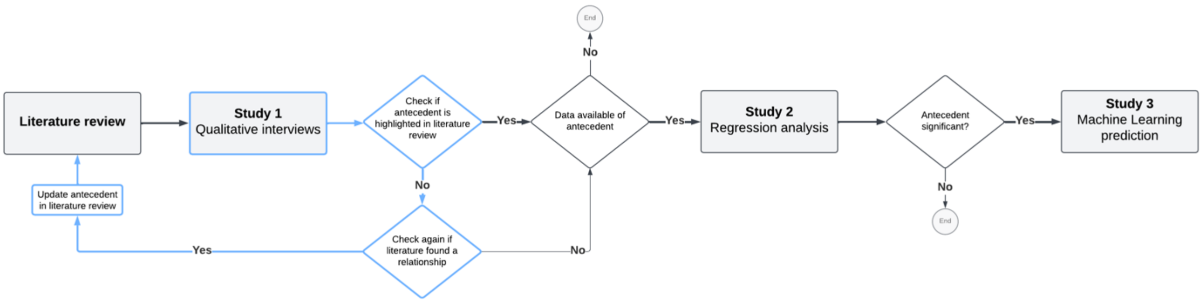

Figure 1 Overview of the Multi-Study Research Approach (Study 1)

This study enables an in-depth exploration of teacher absenteeism and uncovers nuanced factors that quantitative methods may overlook (Brinkmann & Kvale, 2009; Werang et al., 2017). Building on the existing literature, this study affirms known antecedents such as having children, limited autonomy, and additional responsibilities. Moreover, it identifies new antecedents, including employment at multiple schools and unfavourable schedules. Thereby, offering a more comprehensive understanding of teacher absenteeism in Dutch secondary education. Nevertheless, the study finds discrepancies regarding the ambiguity in the directionality of the relationship between the exogenous and endogenous variables (experience, monitoring/inspection). For instance, the literature suggest that absenteeism rates tend to increase with higher levels of experience, indicating a positive relationship between teacher experience and absenteeism. In contrast, qualitative insights reveal that starting teachers are more frequently absent as they lack established resources and experience. Thereby, suggesting a negative relationship. A similar discrepancy arises regarding the influence of monitoring and inspection. Qualitative findings indicate that increased monitoring and inspection contributes to increased stress levels and reduced well-being, ultimately contributing to higher absenteeism. Thus, suggesting a positive relationship. On the other hand, the literature reveals that increased monitoring and inspection reduce absenteeism, hence suggesting a negative relationship. Such differences may arise due to context-specific factors, methodological differences or sample variations. Moreover, many interviewees suggested certain patterns based on their experiences, while this was not quantitatively grounded and lacking empirical validation. Table 3 summarises antecedents of teacher absenteeism as reported in the literature, alongside the antecedents identified in Study 1. The following sections elaborate on the methods used in Study 1 and subsequently describes the results in detail.

Table 3 Antecedents: Literature vs Study 1

	Antecedent	Literature	Study 1
Personal	Age	-	-
	Gender *(female)*	+	+
	Marital status *(married)*	+	
	Education	-	
	Prior absenteeism	+	
	Children	+	+
	Health-status	-	-
	Menopause	+	+
	Perfectionism		+
	Busy social life		+
Job-related	Tenure	+	
	Experience	+	-
	Job-satisfaction	-	-
	Working environment/conditions	-	-
	Working hours	+	+
	Job involvement	-	
	Autonomy	-	-
	Commuting distance	+	
	Stress	+	+
	Leadership style *(supportive)*	-	-
	Additional responsibilities	+	+
	Monitoring/Inspection	-	+
	Managerial absence	+	
	Team-level absence	+	+
	Student behaviour	+	+
	Multiple schools		+
	Digitalisation		+
	Flexibility		-
	Following a study		+
	Less favourable schedule		+
	Hierarchical organisation		+
	Subject *(core subjects)*		+
	School type		+/-
	Municipality		+/-
Socio-economic	Salary	-	
	Contract type *(permanent)*	-	
	Employment status *(full-time)*	-	
	Tight labour market		+
	Tight housing market		+
	Parental influence		+
	Social safety		-
	Public opinion		+

Note: '+' indicates a positive relationship, '-' a negative relationship, '+/-' signifies that interviews suggest a potential impact on absenteeism, though the direction of the relationship remains unclear. An empty cell means that a relationship was not found or mentioned

3.1 METHOD

3.1.1 Research instruments

The primary qualitative research instrument for Study 1 are semi-structured interviews to gather in-depth insights. This approach helps to uncover antecedents that may not be covered in existing literature and thus provides a more comprehensive understanding of the factors influencing teacher absenteeism (Werang et al., 2017). Semi-structured interviews combine the flexibility of unstructured interviews and the systematic approach of structured interviews. It follows a guideline while allowing the exploration of additional topics that emerge during the interview (Brinkmann & Kvale, 2009). The use of semi-structured interviews is supported by examples from existing research. The interview guide can be found in Appendix A.

According to Smith (2015), this approach ensures that all relevant topics are covered while allowing the exploration of additional areas. The interview guide is developed based on previous studies regarding predictors of teacher absenteeism. For example, the study by Green (2014) investigated teacher absenteeism and desire to leave teaching. Also, Werang et al. (2017) presented useful information. Moreover, regular conversations were held with employees at Infotopics, who provide HR analytics to educational institutions. These employees have extensive knowledge in terms of data analytics in the educational sector. Overall, the qualitative findings informs Study 2 where relationships are examined.

3.1.2 Data collection

To select participants for the semi-structured interviews, this study uses a purposive sampling technique. This approach is a non-probability technique in which the researcher selects participants relevant to the study (Palinkas et al., 2015). Interviews were conducted with teachers, (vice) principals, and administrators at Dutch secondary educational organisations. Moreover, multiple conversations with Infotopics employees working in the dedicated department offering analytics to educational institutions were held. According to Baker and Edwards (2012), the amount of interviews during qualitative research depends on the context. They suggest at least eight to ten interviews to draw ground conclusions. To ensure accuracy and verification of statements, the interviews were audio-recorded and transcribed, if consent is given (Deakin & Wakefield, 2013). The interviews lasted approximately 45 minutes and were held in Dutch. Publicly available data was used to gather insights about the organisation and the interviewee, ensuring a thorough preparation before the interview. Table 4 highlights the interviewees of this study.

Table 4 Overview of Qualitative Participants within their Organisations

Company	Function	Organisation size[9]	Gender	Absence rate[10]	Respondent
School A	Principal	± 450	Female	3-5%	R1
School B	Principal and Teacher	± 318	Female	5,2%	R2
School C	Independent expert & researcher	N/A	Male	N/A	R3
School D	Vice principal	± 206	Male	5%	R4
School E	Principal	± 386	Male	2%	R5
School F	Health Management Advisor	± 950	Female	4,6%	R6
School G	Principal	± 197	Female	7,2%	R7
School H	HR Director + Head of HR	± 1500	Male	4,8%	R8

3.1.3 Data analysis

As mentioned, the interviews are transcribed and audio-recorded. This study uses inductive coding, also known as interpretive coding, which identifies relevant factors that can explain a particular phenomenon and identify relationships (Kump & Dahlke, 2024). The inductive coding approach is done through the GIOIA method. This method allows for a systematic approach to organise data into First-order concepts, Second-order themes, and aggregate dimensions to uncover patterns and themes (Gioia et al., 2012; Kump & Dahlke, 2024).
The coding process was done via ATLAS.ti, a software tool for qualitative data analysis[11]. The first step consists out of thematic analysis which involves a process where the researcher transcribes the data, reads it aloud and then re-reads it while recording initial codes (Gioia et al., 2012; Magnani & Gioia, 2023). The transcriptions were created using Artificial Intelligence (AI) tools to have every element on paper. The transcription were adapted manually to avoid errors.

[9] *Number of employees* - DUO. (2023). *DUO open education data*. Retrieved from https://duo.nl/open_onderwijsdata/
[10] Absenteeism rates derived from annual reports or interviews
[11] University of Twente. (2024). *BMS Lab Software Licenses*. https://www.utwente.nl/en/bms/datalab/dataanalysis/software/

The next phase is to develop 'First-order' elements, as proposed by Gioia et al. (2012), in which first codes are generated by highlighting significant data. This method, also known as open coding, ensures that every statement is involved (Kump & Dahlke, 2024). This step led to 242 initial codes and were migrated to an Excel. Initial codes were grouped into categories and labelled, resulting in 79 First-order concepts.

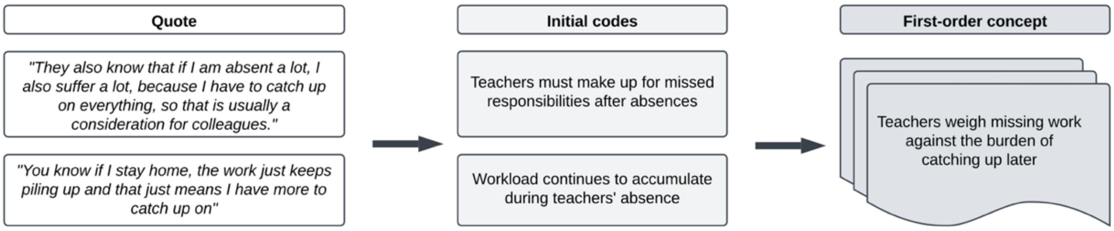

Figure 2 Demonstration of First-order concept formation

In the subsequent phase, 'Second-order themes' were developed by looking for possible themes. The initial codes and First-order concepts created in the previous step should support the Second-order theme. This step is also known as axial coding, in which the researcher is considered as a 'knowledgeable agent' (Gioia et al., 2012). In total, 13 Second-order concepts were created.

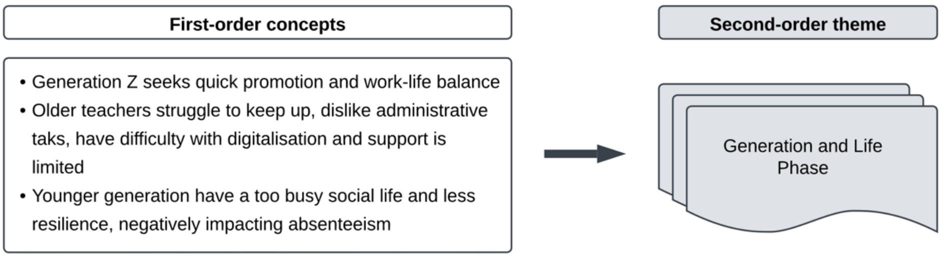

Figure 3 Demonstration of Second-order concept formation

The ultimate stage is to create aggregate dimensions based on Second-order concepts. Finally, a thematic map is created based on first- and second-order concepts and aggregate dimensions (Kump & Dahlke, 2024; Magnani & Gioia, 2023).

3.2 RESULTS

This paragraph presents and elaborates on the concepts, themes, and dimensions that emerged from the qualitative data analysis. For consistency throughout this study, the GIOIA structure remains to classify categories of absenteeism for the aggregate dimension from Porter and Steers (1973). These include: personal, job-related, and socio-economic antecedents. The GIOIA structures are illustrated in Figure 4, 5, and 6.

3.2.1 Personal antecedents

The first central dimension, personal antecedents, contains individual characteristics and personal circumstances influencing teacher absenteeism. Through in-dept interviews, two main Second-order themes emerged: **health and well-being**, and **generations and life phase**. These factors reflect the impact of personal health, well-being, family responsibilities, and generational life phases

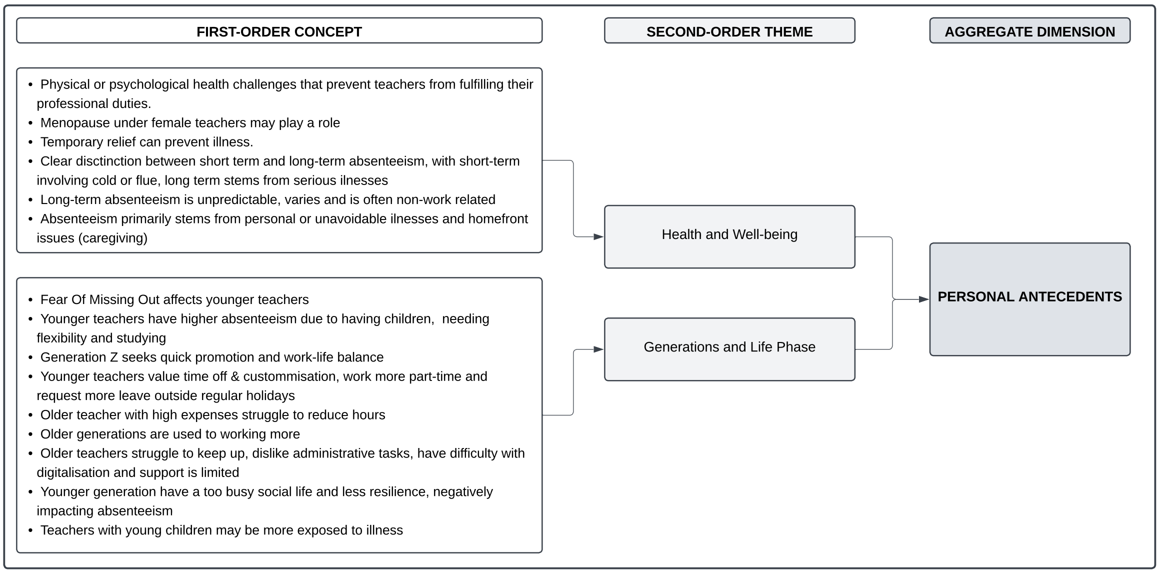

Figure 4 GIOIA Structure for Personal Antecedents

Health and Well-being

Health challenges, both physical and psychological, were repeatedly highlighted as major drivers: *"The absence due to mental or physical complaints, which prevents the colleague from fulfilling the obligations that may be required (R1)"*. Another emphasised, *"Absence caused by physical or psychological complaints, specifically absence due to illness (R7)"*. Moreover, it was mentioned that gender-related issue, such as menopause, appears influence absenteeism among female teachers: *"Women approaching menopause do experience more* *complaints […] but I notice that it is indeed a factor for some women. I don't have exact figures on how this specifically applies to us (R4)"*. While another stated: *"But also the situation or context someone is in […] menopause (R7)"*. This issue is often overlooked, while having an impact on well-being and attendance.

Furthermore, it was highlighted that absenteeism primarily stems from personal or unavoidable illnesses and home front issues, such as caregiving: "*Much of the absenteeism I observe has a private cause. The people currently absent are dealing with issues the school cannot address. Hospital treatments and caregiving responsibilities are actually the main causes of absenteeism (R5)"*. This aligns with participants' observations that absenteeism is unpredictable, variable, and often non-work related: *"Unfortunate circumstances with some colleagues who are indeed on long-term leave for non-work-related reasons. Someone who fell down the stairs, someone who suddenly had an epileptic seizure […] these incidents are not directly related to school or their work (R4)"*. Another added: *"Significant peaks and also some troughs, but the largest peaks far exceed those (R7)"*. A clear distinction between short- and long-term absenteeism was noted. With short-term involving a flu or cold and long-term serious illnesses: *"There is indeed a clear distinction between short-term and long-term absenteeism within our organization (R2)"* and *"Short-term absenteeism involves occasional days off to recharge or a few days of mild illness, while long-term absenteeism is due to issues such as burnout symptoms, accidents, or serious illnesses (R4)"*.

Generations and Life Phase

The second theme emerging from the interviews explores how teachers of different ages face unique challenges that are shaped by generational differences, life-phases, and personal responsibilities. First of all, younger teachers, often in the early stages of their careers, struggle to balance family responsibilities alongside their professional obligations: *"In the middle group, aged 25-40, children and studies play a significant role, taking up a lot of their time (R6)"*. Others also emphasised: *" Young children do have an impact, by the way"(R2),* and *"[…] whether or not they have children (R4)"*. This highlights that teachers with childcare responsibilities exhibit higher absenteeism rates due to the greater flexibility these demands require. Moreover, teachers with young children are more exposed to illness: *"If you have very young children who go to daycare […] they are exposed to significantly more viruses and bacteria (R2)"*.

Generational differences also play a role in absenteeism patterns. Younger teachers, particularly Generation Z, emphasise quick career progression and work-life balance. One respondent explained: *"[…] they tend to be less patient and expect to advance quickly in their careers. While in the past one needed considerable patience to climb the career ladder, people now have higher expectations. When promotions don't happen quickly enough, I see that individuals become more dissatisfied, and dissatisfied people are more likely to be absent (R5)".* This generation also values customised work arrangements and additional time off to accommodate their dynamic life demands: *"In the younger generation, I notice that they are more inclined to take breaks to enjoy life more […] they also seek more customised solutions (R4)".* Thus, unmet expectations may contribute to absenteeism. On top of that, participants linked absenteeism to overly busy social lives and reduced resilience: *"And also the social life, which always has to be fun and exciting, with weekends fully packed, often leads to the choice to let work be the first thing to drop (R6)".* Another participant stated: *"The recovery capacity, particularly among Generation Z and Y, is lower due to their busy personal lives (R1)"*. These insights suggest that younger teachers may demonstrate higher absenteeism rates compared to other generations.

Conversely, older teachers face challenges in adapting to new administrative responsibilities and technological changes: *"There are also colleagues approaching retirement who think, 'Do I really want to get involved in this?' They have no interest in it at all (R2)".* This aligns with another respondent: *"Older employees struggle to continue doing what they do. You can see that those aged 55-60 and above find it challenging to keep up and have difficulty adapting to all the changes (R6)".* Moreover, financial pressures also play a role, as older teachers often cannot reduce their working hours due to high expenses: *"In the older generation, you often see that when they are in a phase where their children are studying and their house is not yet fully paid off, there is a combination of wanting to work less but being unable to because the costs are still too high (R4)".*

Thus, these insights indicate that absenteeism is influenced by various personal antecedents: physical and psychological health, family responsibilities (e.g., children and caregiving), gender-related issues (e.g., menopause), and generational differences. Younger teachers often face higher absenteeism due to balancing childcare, exposure to illnesses, and unmet expectations for flexibility and career progression. In contrast, older teachers struggle with adapting to technological changes and administrative demands.

3.2.2 Job-related antecedents

The second central dimension, job-related antecedents, highlights the influence of dynamics, workload, and organisational factors. These factors reflect organisational challenges within schools. Eight themes are included in this dimension.

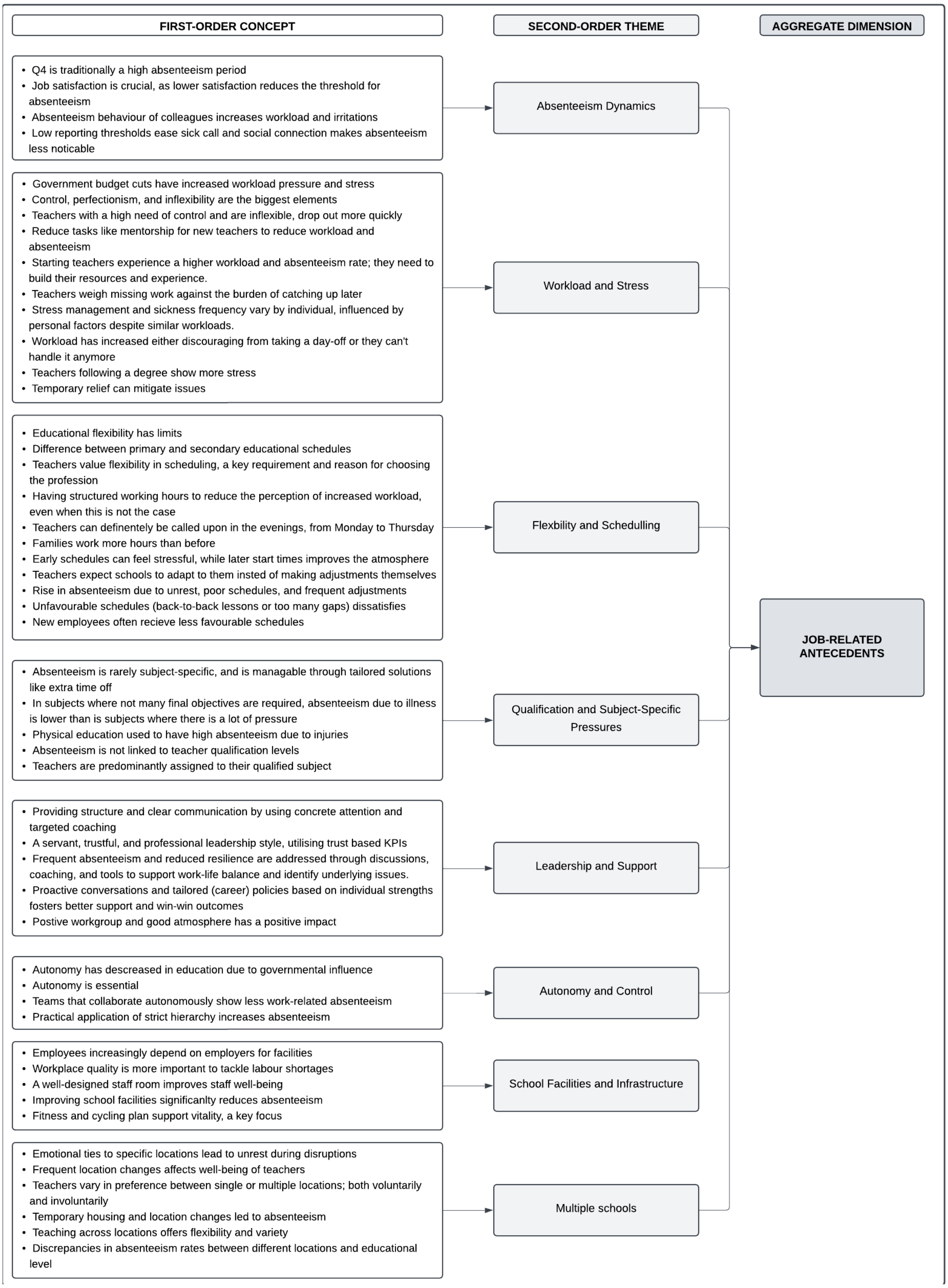

Figure 5 GIOIA Structure for Job-related Antecedents

Absenteeism Dynamics

The first theme of job-related antecedents is absenteeism dynamics. Teachers with lower job satisfaction are more likely to be absent, as dissatisfaction lowers the threshold for taking time off: *"If you no longer enjoy your work and you're nearing the end, the threshold to say, well, I don't feel good enough, is lower (R2)"*. Another participant observed: *"Whether you enjoy the work or have nice colleagues is also a major factor in whether or not you show up (R6)"*. Although participants did not identify an explicit absenteeism culture, they highlighted that peers' absenteeism increases workload. One noted: *"In the French department, I have someone who is on long-term leave […]. The French department is clearly expressing that they're on the verge of collapsing because the workload is so high. With just three of them, they're trying to cover a gap of 1.0 FTE, and that's simply not manageable (R1)"*. Another emphasised that it creates irritations: *"But you do notice that this quickly leads to irritation. People who frequently call in sick for short periods are not always well received within a department, as it tends to create a sense of judgment (R6)"*. This underscores the link between team dynamics and individual attendance. Furthermore, a recurring theme is the seasonal trend in absenteeism, with Q4 being a period of increased absences: *"Traditionally, Q4 is the weakest period (R5)"* and *"In this period we're in now, naturally, it's always more (R4)"*. Additionally, low reporting thresholds and social connections make absenteeism less noticeable and more accessible for individuals to adopt: *"One of the main reasons is that we have a very low threshold for reporting absences. The process is very easy, almost without barriers—people simply send an email. The whole social context, the feeling of being missed, has diminished in my view. It makes it easier to stay away for a day because it goes unnoticed (R7)"*.

Workload and Stress

Workload and stress were identified as major antecedents driving absenteeism. Due to government budget cuts, workload and stress increased: *"Education is facing budget cuts […] they have to do the work with fewer and fewer people, so the workload is quite high. That doesn't always help (R6)"*. Moreover, teachers with a high need of control, perfectionism and inflexibility are more prone to drop out: *"Control tendencies, […] perfectionism […], lack of flexibility […] are the three main elements […] those dropout more quickly (R5)"*. New teachers particularly face high workloads, as they lack established resources and experience, leading to increased absenteeism: *"But that's not without reason, because especially people who are just starting out in education—whether they are* *newcomers or fresh out of training—experience a high workload. In the beginning, you have to invest a lot because you have nothing prepared yet (R2)"*. Another participant remarked: *"Currently, it is indeed noticeable (starting teachers have higher absenteeism) (R6)"*. Although initiatives are in place to reduce workload for new teachers- allocating 20% less work in their first year and 10% less in their second. Nevertheless, due to the tight labour market, these measures are always feasible: *"That 10/20% reduction and all the support that starters are supposed to receive does not always happen at the moment, precisely because there is often a shortage of staff (R6)"*. To address absenteeism in this group, one suggested to reduce additional responsibilities (e.g. mentorship tasks): *"If you want to prevent starting teachers from taking sick leave, you need to reduce their workload [...] you have to be cautious with assigning school tasks (R5)"*

Over the last years, work pressure has increased among teachers: *"I believe the workload has truly increased and is, on average, far too high (R3)"*. Teachers often weigh the decision to take a day off against the burden of catching up on missed work later: *"You know, if I stay home, the work just keeps piling up, which only means I will have more to catch up on later (R2)"*. This sense of responsibility can lead teachers to push themselves beyond their limits, as another participant noted: *"The sense of responsibility that leads teachers to push themselves too far because they still want to give that lesson (R7)"*.

The experience of stress and absence management is highly individual, even with similar workloads: *"Some people experience stress much sooner than others, and this first group generally tends to call in sick more frequently (R4)"*. Simultaneously, another pointed out: *"I really have to admit that one colleague stays home with a cold, while another still comes to school even with a throat infection (R2)"*. Finally, teachers pursuing further qualifications may also face increased workload, contributing to absenteeism: *"People who are still in teacher training, that does play a role (R6)"*.

Flexibility and Scheduling

Flexibility and scheduling are critical factors influencing absenteeism, as they directly impact job satisfaction and workload perception. Teachers value flexibility in their schedules, which is often a key reason for choosing the profession: *"The colleagues are at school when the lessons are being taught, but they can decide for themselves when they prepare their lessons and carry out their tasks (R1)"*. At the same time, another pointed out: *"People deliberately* *choose to teach in secondary education because, for example, in the afternoon they can say the children don't need to go to daycare (R2)"*. However, the flexibility of teaching schedules has limits: *"The reality of education offers significant flexibility, but it has its limits. It cannot be fully adapted to meet the preferences of every individual teacher, as this would result in an unworkable schedule or one that is highly disadvantageous for students (R4)"*. It emphasises that teachers expect schools to adapt to their needs while neglecting students' schedules.

Teachers often have the perception of a high workload, but due to the flexibility, actual workloads remain unchanged: *"(workload) is a matter of perception; when we add up the hours, it amounts to a standard 40-hour workweek (R7)"*. At the same time, another pointed out: "*We have one school where teachers are really on-site from 8 to 5. […] However, they feel that when they go home, they have accomplished more and can truly relax. […] I observe very little work-related absenteeism there (R6)"*. Participant highlighted that having clear and predictable schedules might improve workload perception: *"Wouldn't it be much better for education to establish that everyone arrives at 8 a.m., everyone leaves at 5 p.m., and the evenings are free (R1)"*

Early schedules can also feel stressful, starting later in the morning significantly improves the atmosphere: *"If you want to create a good atmosphere in your school, you should start a bit later […] that helps tremendously (R5)"*. Similarly, another brought to the attention that it creates stress: *"[…] They do feel that not having the first hour free brings a certain pressure or stress with it (R4)"*. This aligns with a recurring theme that poor schedules, frequent adjustment, and temporary locations creates dissatisfaction and impacts absenteeism: *"The unrest, temporary housing, having to adjust more, and less favourable schedules […] you could see absenteeism increasing (R4)"* and *"Changes in location and favourable or less favourable schedules (R3)"*. Poor schedules include starting early or finishing late and having too many gaps between lectures: *"A schedule with many gaps, lessons at the end of the day, and situations where people's preferences cannot be accommodated […] that is considered an unfavourable schedule (R6)"*. This dissatisfaction leads to a lower threshold for absenteeism: *"This leads to greater dissatisfaction and, ultimately, a lower threshold for absenteeism (R6)"*. New employees are particularly vulnerable, often receiving less favourable schedules, which adds to their workload and stress: "*It is often a common reflection that, as a new employee, you are given a less favourable schedule (R3)"*.

Qualification and Subject-Specific Pressures

According to participants, specific subjects contribute relatively minor to absenteeism compared to other antecedents. Absenteeism is rarely tied to specific subjects: *"With us, there isn't a specific subject that is more frequently cancelled (R2)"*. Tailored solutions, such as granting extra time off, were effective in managing subject-specific absenteeism: *"It's not subject-related, and where it is subject-related, absenteeism can often be prevented by offering tailored solutions [...] such as an extra free afternoon (R5)"*. Nevertheless, differences were observed based on core subjects with higher demands. For instance, mathematics and Dutch were noted to experience higher absenteeism rates: *"In particular, subjects that do not require homework and those with fewer final exam requirements tend to have lower absenteeism rates compared to subjects with higher pressure. Subjects such as English, mathematics, Dutch, French, German, physics, and chemistry experience higher absenteeism rates (R1)"*. This highlights discrepancies between schools. Moreover, some indicated that analysis at this level has not been conducted; thus this cannot be scientifically concluded.

Physical education, used to have higher absenteeism rates due to injuries: *"In the past, [...] there was much more absenteeism in subjects like physical education, for example, due to injuries (R1)"* and *"You also sometimes find that physical education teachers develop physical complaints over time, though these are not necessarily work-related (R2)"*. Additionally, absenteeism is not linked to teacher qualification levels: *"No difference in absenteeism between first-degree and second-degree teaching qualifications (R1)".* Another also illustrated: *"No, it's not the case that first-degree teachers are more or less frequently sick than second-degree teachers (R2)"*. Moreover, teachers are predominantly assigned to their qualified subject, with any non-qualified subjects being voluntary: *"Rarely, when we have understaffing and ask colleagues for help, it is largely on a voluntary basis (R4)"*. Similarly, others stated: *"I think it's going to become a trend [...] but [...] at the moment, they are assigned to the subject they teach 99% of the time (R5)"* and *"That willingness is generally very high, though the risk is that people sometimes demand too much of themselves. We need to help maintain that balance at times (R4)"*. This suggests, due to a tight labour market, the trend of assigning teachers outside their qualifications, even on a voluntary basis, could influence absenteeism if a balance is not carefully maintained.

Leadership and Support

Leadership and support plays a role in reducing absenteeism by providing structure and clear communication through concrete attention and targeted coaching: *"Structure and clarity, so a lot of communication […], we have different types of coaches […] to help where the need for support lies and to assist in discovering that (R1)"*. This approach ensures that specific needs and/or issues are identified and addressed effectively. Moreover, a servant, trustful, and professional leadership style that utilises trust-based Key Performance Indicator (KPIs) is used by management teams to reduce absenteeism: *"Trust as a fundamental attitude and allowing space for professionalism and autonomy greatly benefits the school environment. Hard KPIs are quite different in education, so the second, more dialogical approach is better suited (R5)"*. This balance was further elaborated: *"It should not be too strict, but also not too lenient (R4)"*. Additionally, tailored career solutions further enhance support by addressing individual needs, resulting in mutual benefits: *"An open conversation to see what comes out of it and how we can help someone—it's usually a win-win situation (R6)"*. There has been a shift from passive to proactive discussion to explore career trajectories and alternative potential options: *"In the past, the approach was more passive, but now we take the initiative more quickly to engage in conversations […] discussing other options, such as possibly separating earlier or pursuing a career development trajectory (R6)"*. These findings underscore the importance of a servant leadership approach that fosters communication, trust, and individualised support.

Autonomy and Control

Autonomy and control have emerged as essential themes in maintaining professional satisfaction and effective teaching practices: *"Autonomy is incredibly important for everyone, especially for teachers (R4)"*. Nevertheless, autonomy has decreased due to increasing external pressures: *"The fact that autonomy has decreased (R1)"*. Government policies dictate work routines, reducing teachers' ability to operate independently: *"How problematic is it that part of your workday is dictated by government directives? We are required to allocate a certain percentage of our teaching time to four core skills. Nobody asked for that (R1)"*. This decline in autonomy is also linked to increased workload: *"Yes, I do think it has decreased […] you have to do the same work with fewer and fewer people (R6)"*. The positive effects of autonomy at team level was also emphasized. Collaborative and self-managed teams demonstrated reduced absenteeism rates, indicating a supportive work environment: *"They leave it entirely up to the team, which is willing to go the extra mile and* *help one another. Yes, that approach works well. I see very little work-related absenteeism there (R6)"*. This aligns with the practical application of strict hierarchical structures: *"On paper, we are a hierarchical organization, but if you implement that in practice, it leads to high absenteeism (R5)"*. Suggesting that rigid control mechanism reduces autonomy mechanisms, ultimately contributing to higher absenteeism rates.

School Facilities and Infrastructure

The next theme that emerged is school facilitates and infrastructure. Schools with improved infrastructure reported lower absenteeism rates, as highlighted by one school with an absence rate of just 2%. Increasingly, employees rely on employers for resources and support: *"As an employer, you need to do more to support your employees. The employee [...] becomes somewhat more dependent on the employer in a facilitative sense (R5)"*. Suggesting the growing responsibility of employers to provide adequate resources and support.

The quality of the workplace environment is crucial for addressing labour shortages and enhancing well-being: *"Higher standards are being placed on the workspaces due to the tight labour market [...] colleagues are quite attentive to this. We are working on improving the facilities at school, and we've noticed that absenteeism is significantly decreasing (R5)"*. Access to tools and materials for effective job performance also plays a role: *"Work environment, to what extent do you have access to the appropriate materials (R2)"*. Programs promoting employee vitality, such as fitness and cycling initiatives, were also mentioned as beneficial: *"Vitality is a relevant theme (R4)"*. These findings highlight the importance of investing in workplace facilities and infrastructure to reduce absenteeism and promote well-being.

Multiple schools

The last job-related theme that emerged is working at multiple schools. Teachers often develop strong emotional ties to specific locations, and disruptions can cause unrest: *"I deliberately chose this location, or I have given my heart to this location [...] so it causes a lot of urns (R1)"*. Suggesting that location changes can create stress and dissatisfaction. Preferences for working at single or multiple locations vary among teachers– both voluntarily and involuntarily– depending on personal and professional circumstances. For some, multiple locations offers flexibility, variety, and contract hours: *"I also have a colleague who says: I actually really enjoy the variety (R2)"* and *"If you want more hours and I can't offer them to you [...] then you can work at another school (R7)"*. While others, emphasise that multiple locations increase workload due to more meetings: *"That reduces meetings [...] and discussions (R2)"*. Moreover, it was emphasised that location in which the school is located (village or town) and school type might influence absenteeism: *"We suspect that absenteeism might vary based on school type or whether the school is located in a village or town (R8)"*.

Working at multiple locations, were associated with stress, dissatisfaction, and reduced well-being, ultimately contributing to higher absenteeism: *" We have reduced from three to two locations, and I think that greatly benefits people's well-being. Operating across multiple locations works against you (R5)"* and *"Yes, I am absolutely certain of that" (frequent switching between locations impacts well-being and absenteeism) (R6)"*. Moreover, another participant highlighted: *"We just finished a renovation, and we noticed that absenteeism really increased during that time. Instead of being able to take a moment to rest during a free period, you had to spend it relocating (R4)"*. Thus, these disruptions result in lost time between lessons, ultimately affecting teachers' stress levels and workload. Suggesting that frequent location changes can ultimately lead to higher absenteeism rates. Additionally, it might be that absenteeism differs between municipality and school type.

3.2.3 Socio-economic antecedents

The final dimension, socio-economic antecedents, reflects **labour market pressures**, **financial challenges**, and **external pressures** that influence absenteeism.

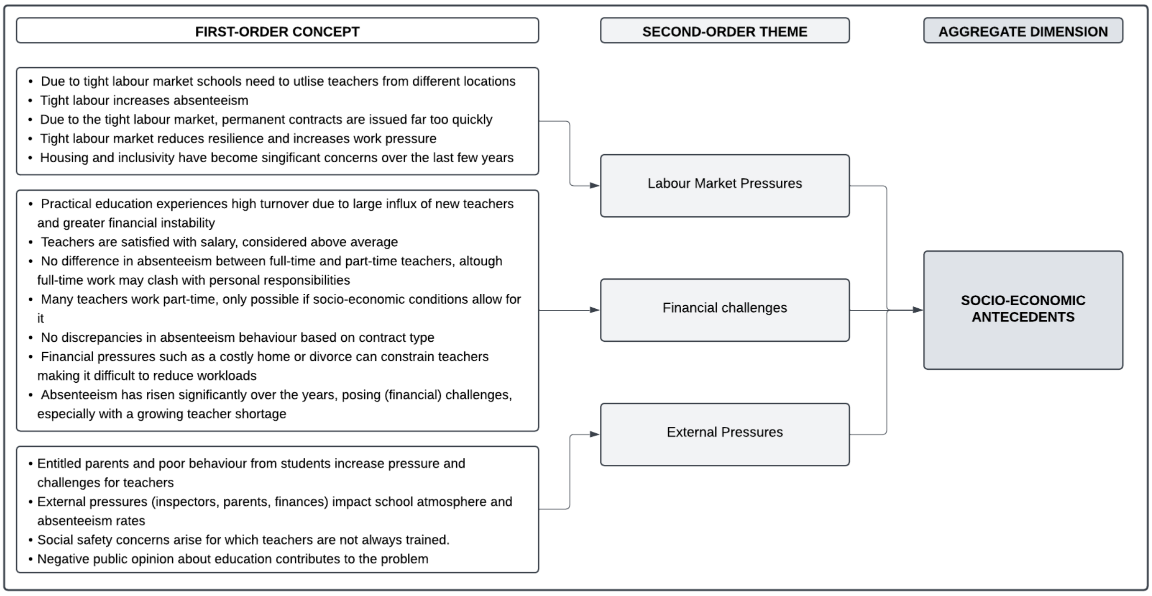

Figure 6 GIOIA Structure for Socio-Economic Antecedents

Labour Market Pressures

The first socio-economic theme, labour market pressures, significantly influences workload, well-being, and teacher absenteeism. The current tight labour market forces schools to assign teachers from different locations to address staff shortages: *"Mandatory because, due to the tight labour market, there are no substitute teachers available, so a teacher from another location has to step in to ensure education can continue at the location where the teacher is absent. This, in turn, leads to significant dissatisfaction (R1)"*. The shortages also directly influences workloads for remaining teachers, particularly in cases of long-term absenteeism: *"In the French department, I have someone on long-term leave […] the French department is clearly expressing that they're on the verge of collapsing because the workload is so high. With just three of them, they're trying to cover a gap of 1 FTE, and they're saying it's not manageable (R1)"*. A consequence is the quick issuance of permanent contracts, which can result in teachers being in roles that are not well-suited to them: *"I believe that due to the tight labour market, permanent contracts are being issued far too quickly, resulting in colleagues being placed in unsuitable positions. However, because of the "golden cage" of the education sector's collective labour agreement, they remain in these positions, ending up in a workplace that is not a good fit for them (R2)"*. These mismatches have long-term implications for teacher satisfaction, retention, and absenteeism. Over the last two years, the tight housing market has also emerged as a factor, reflecting broader societal challenges that `make challenging decisions, such as quick contract issuance and teacher relocation, that may lead to dissatisfaction and absenteeism.

Financial Challenges

Financial challenges emerged as another theme, highlighting external issues and pressures. First of all, practical education faces higher turnover rates due to a large influx of new teachers and financial instability, as one noted: *"The unrest in education regarding finances affects practical education, particularly with the practical classrooms. Practical education [...] is dealing with a relatively large proportion of new teaching staff due to many departing teachers, and the cost of onboarding new teachers is comparatively high (R1)"*. Such financial uncertainties disrupt stability and creates unrest.

Teacher salary is generally considered above average in Dutch secondary education, thereby not influencing absenteeism: *"Teachers generally earn a decent salary, so overall, that aspect is fairly well addressed (R4)"*. Moreover, many teachers work part-time, which is only viable under stable socio-economic conditions: *"Teachers generally have a decent salary, so overall, that aspect is fairly well-covered (R2)"*. Additionally, no difference was found between full-time and part-time teachers in terms of absenteeism. Nevertheless, it was briefly mentioned that full-time work might create stress when it clashes with personal obligation (e.g., caregiving responsibilities) *"You can't work full-time and simultaneously say, "I'm not fully available to work"; that's obviously a strange combination (R4)"*.

Furthermore, most participants did not find a relationship between contract type and absenteeism: *"No, it is not the case that people with permanent contracts are more likely to be absent (R2)"* Only one noted that temporary contract can create stress due to job insecurity. Broader financial pressures, such as costly housing or divorce, further influence stress and absenteeism. Participants emphasised that absenteeism has risen significantly in recent years, due to growing teacher shortage, disrupting school operations: *"That is, of course, quite a problem in schools, especially with the growing teacher shortage. On a school with around 130 teachers, this means you have 6-7 people absent every day [...] which is quite a lot (R4)"*.

External Pressures

The last socio-economic theme that emerged is external pressures. Recurring antecedents are the increasing pressures caused by entitled parents and student behaviour. Participants highlighted that some parents assume an authoritative role over teaching practices, which complicates the profession: *"What is very apparent at my location is that we have a profession where parents often believe they have the same expertise (R1)"*. This, combined with challenging student behaviour: *"Due to the behaviour of students—and not to forget, parents—face significant challenges (R6)"*. Due to assertiveness of parents, teachers have to maintain strong resilience: *"I also think that outspoken parents play a role as well—you need to stand your ground [...] which makes it increasingly challenging (R7)"*. Moreover, the school environment and the intensity of student behaviour were linked to absenteeism: *"However, the context of the school and the varying personalities of the children [...] the more restless or demanding the children, the higher absenteeism can become (R7)"*. Beyond parents and student behaviour, broader external pressures were seen to negatively impact the overall school atmosphere and absenteeism rates: *"You notice that when there is external pressure on a school [...] across various factors—inspection, parents, finances—and as management, you have to take a very directive approach, it directly impacts the atmosphere and is immediately reflected in absenteeism rates (R5)"*. This demonstrates how external pressures directly influence workplace morale and teacher attendance.

Furthermore, concerns surrounding social safety have also emerged. Teachers are increasingly expected to handle issue such as criminal or inappropriate behaviour, which goes beyond their professional focus: *"Currently, a significant focus is on social safety [...] dealing with criminal behaviour and unacceptable conduct, for which teachers are not always trained. Teachers are often more focused on the content of their subjects, but now they must manage students who are outright rude and aggressive. If teachers are not equipped to handle this, you do see people stepping away (R6)"*. This contributes to teacher stress, disengagement, and potentially leaving the profession. Lastly, the negative public opinion surrounding education was emphasised as a factor. *"Everyone thinks they know better about education. I truly believe that the ongoing tendency to talk about education—often in a negative light—has not improved. I think this has also played a significant role (R7)"*. These findings underscore the significant impact of external pressures on teacher absenteeism, including entitled parents, challenging student behaviour, social safety concerns, and negative public opinions. Suggesting that such antecedents may lead to a demanding and unsupportive environment, contributing to increased stress, dissatisfaction, and absenteeism rates.

4. STUDY 2 - REGRESSION ANALYSIS

Building on the findings from the literature review and Study 1, Study 2 (highlighted in blue) quantitatively tests the antecedents of teacher absenteeism using regression analysis from two dependent variables: absence duration and absence frequency. These measures are commonly used in absenteeism studies (Gellatly, 1995; Knoster, 2016; Porter & Steers, 1973; Rosenblatt & Shirom, 2005). Figure 7 presents the overall multi-study approach adopted in this research. In Study 3, the significant antecedents identified in this study are used as model features in Machine Learning models to predict absenteeism categories.

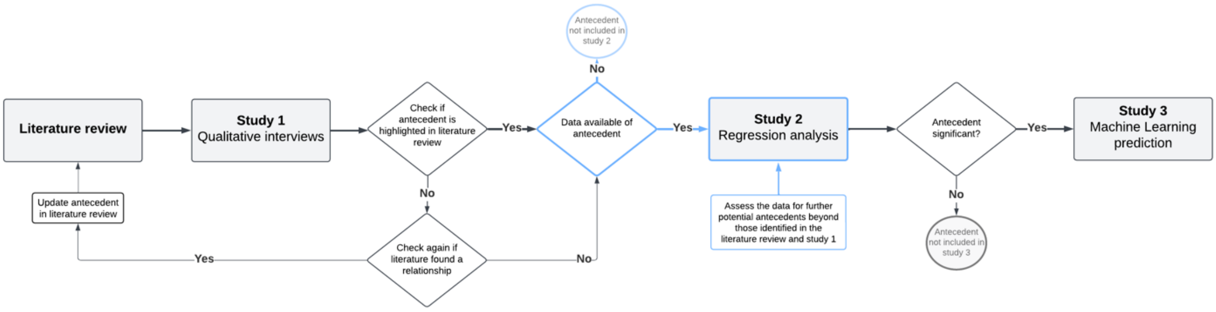

Figure 7 Overview of the Multi-Study Research Approach (Study 2)

Utilising the classification of antecedents influencing absenteeism from Porter and Steers (1973), which categorises absenteeism into personal, job-related, and socio-economic antecedents, Table 5 presents the hypotheses tested in this study. These are based on the identified antecedents in the literature review and Study 1. In addition, this study examines data-driven antecedents (age of manager, replacement/ temporary vacancies, FTE in team, and function), as relevant data were available. Unfortunately, not all the antecedents that are mentioned in the literature and interviews were available.

The available antecedents are tested using a comprehensive HR dataset from teachers within a Dutch secondary educational institution. For some variables, no clear directional hypotheses could be formed due to limited or inconclusive prior research. In such cases, non-directional hypotheses are formulated, expecting that a relationship is expected, though the direction remains uncertain. This study offers a comprehensive and grounded perspective on the multifaceted antecedents of teacher absenteeism. First, it contributes to teacher absenteeism literature by confirming, refinining, and rejecting established antecedents. (Harrison & Martocchio, 1998; Knoster, 2016; Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991), thereby strengthening or questioning existing assumptions.

Second, it identifies novel antecedents (FTE in team, age of manager, school type, and employment at multiple schools) that have received no attention in prior literature. The following sections outline the hypotheses development, methodological approach, data description, and describes the results in detail.

Table 5 Hypotheses Development

	Antecedent	**Literature**	**Study 1**	**Study 2 Hypotheses**	
Personal	Age	-	-	H1	-
	Gender *(female)*	+	+	H2	+
	Marital status *(married)*	+			
	Education	-			
	Prior absenteeism	+		H3	+
	Children	+	+		
	Health-status	-	-		
	Menopause	+	+	H4	+
	Perfectionism		+		
	Busy social life		+		
Job-related	Tenure	+		H5	+
	Working hours (FTE)	+	+	H6	+
	Function			H7	+/-
	Experience	+	-		
	Job-satisfaction	-	-		
	Working environment/conditions	-	-		
	Job involvement	-			
	Autonomy	-	-		
	Commuting distance	+			
	Stress	+	+		
	Leadership style *(supportive)*	-	-		
	Additional responsibilities	+	+		
	Monitoring/Inspection	-	+		
	Age of manager			H8	+/-
	Managerial absence	+		H9	+
	Multiple schools		+	H10	+
	FTE in team			H11	+/-
	Team-level absence	+	+	H12	+
	Student behaviour	+	+		
	Digitalisation		+		
	Flexibility		-		
	Following a study				
	Less favourable schedule		+		
	Hierarchical organisation		+		
	Subject		+/-		

Table 5 Cont.

	Antecedent	Literature	Study 1	Study 2 Hypotheses	
Job-related	Municipality		+/-	H13	+/-
	Replacement/temporary vacancies			H14	+/-
	School type			H15	+/-
Socio-economic	Salary	-		H16	-
	Contract type *(permanent)*	-		H17	-
	Employment status *(full-time)*	-		H18	-
	Tight labour market		+		
	Tight housing market		+		
	Parental influence		+		
	Social safety		-		
	Public opinion		+		

Note: '+' indicates a positive relationship, '-' a negative relationship, and '+/-' means a non-directional relationship, implying a difference is expected though the direction of the relationship remains unclear. An empty cell in the columns 'Literature' and 'Study 1' means that no relationship was found. An empty cell in the column 'Study 2 Hypotheses' means that no data was available to test the relationship.

4.1 METHOD

4.1.1 Research instruments

This research is conducted in collaboration with Infotopics B.V. The organisation, established in 2003, offers Business Intelligence (BI) and data visualisation solutions in four areas: 1) consultancy, 2) software licenses for data visualisation (e.g. Tableau, Alteryx, and Snowflake), 3) providing an academy for these software programs, and 4) sector solutions beyond standard functionalities[12]. The data collection instrument that is used for the quantitative part is HR data provided by Infotopics in PowerBI[13]. Infotopics has a dedicated business unit focused on the education sector. providing specialised educational analytics by offering comprehensive HR dashboards to educational intuitions, which include absenteeism, performance metrics and other relevant HR variables While extensive data of HR analytics from a Dutch secondary education is available, advanced analytics are not included. In practice, Infotopics frequently receives inquiries from clients in the education sector to gain better insights into teacher absenteeism. This data serves as the primary source of quantitative information for this study. Using existing data can lead to more efficient research since it reduces time and costs (Johnston, 2014). However, existing data might be missing or incomplete (Smith, 2008).

[12] Infotopics. (2023). *Infotopics DNA - Haal waarde uit data*. https://infotopics.nl/over-ons/
[13] Power BI is Microsoft's platform for creating and sharing interactive data analytics

4.1.2 Descriptive statistics

Study 2 uses existing HR data over the academic years 2022-2024, provided by Infotopics in PowerBI, from a Dutch educational institution. The initial dataset contained 1288 teachers; however, 40 duplicate values were identified. These represented individuals who transitioned from teacher in training to teachers within the same schoolyear. Therefore, the most recent entry for each duplicate was kept. Additionally, one individual with an FTE of zero was removed. Moreover, seven teachers did not have a manager assigned and interns were removed. Consequently, the final dataset consisted of 1150 individuals. Based on the data, a descriptive statistics overview is highlighted in the Table 6, providing key characteristics and distributions (Field, 2018).

Table 6 Descriptive Statistics

Antecedents	Description	Quantity	Mean	Min.	Max.
Absence duration	2023/2024		25.12	0	365
Absence frequency	2023/2024		1.2	0	11
Gender	Male (1)	495			
	Female (0)	655			
Age	< 30	146	44	22	78
	30 – 39	310			
	40 – 49	266			
	50 – 59	215			
	60 +	213			
Function	Teacher LB	719			
	Teacher LC	250			
	Teacher LD	181			
Salary scale	LB (3)	719			
	LC (4)	250			
	LD (5)	181			
Working hours (FTE)			0.71	0.01	1.12
FTE in team			22.2	4.36	38.3
Employment status	Part-time (0)	921			
	Full-time (1)	229			
Menopause	Yes (1)	154			
	No (0)	996			
Age of manager			47.6	33	63
Managerial absence (duration)	Days		6.2	0	162
Managerial absence (frequency)			0.3	0	3
Prior absenteeism (duration)	2022/2023		19.9	0	365
Prior absenteeism (frequency)	2022/2023		0.9	0	7
Multiple schools	Yes (1)	43			
	No (0)	1107			
Team-level absence (duration)			1055	4	3342
Team-level absence (frequency)			35.6	1	79
Tenure	Years		6.4	0.7	30
Contract type	Fixed (1)	661			
	Permanent (0)	489			
Municipality	Town (1)	575			
	Village (0)	575			
Replacement/Temporary vacancy	Yes (1)	512			
	No (0)	638			
Absence duration category	Short-term	787			
	Medium	250			
	Long-term	113			
Total *(n)*		**1150**			

4.1.2 Data analysis

First of all, PowerBI and R studio were used to transform data and test the relationship between the exogeneous variables and endogenous variable due to its flexibility, statistical computing functions and support for data visualisation (Edge, 2019; Kolk, 2021). As mentioned, this study conceptualises the endogenous variable teacher absenteeism 'as any period during which a teacher is absent from their scheduled duties, excluding planned maternity or pregnancy leave'. The variable is operationalised using HR data captured at individual level, provided by Infotopics. Absenteeism was analysed from two dependent variables, the aggregated individual absence duration and aggregated individual absence frequency. These measures are commonly used in absenteeism studies (Gellatly, 1995; Knoster, 2016; Porter & Steers, 1973; Rosenblatt & Shirom, 2005). Absence frequency represents the total number of absence occurrences recorded per individual, whereas absence duration represents the total amount of absence calendar days. It is expected that employees with higher FTE potentially experience higher absenteeism rates due to longer and more weekly work-days (Judge et al., 1997; Rosenblatt & Shirom, 2005). Therefore, the total absence calendar days recorded from the first day of notification to the recovery data per individual is utilised in this study, as specified by the Dutch absenteeism framework (NVS) [14]. Thus, not limiting the analysis to working days of a given period (e.g. 365 days absence duration, while working 0,6 FTE), which prevents bias.

For this study, the 2023/2024 school year is used, instead of calendar year, aligning with the academic calendar. This ensures consistency with term structures and holidays while minimising disruptions between two academic years. Additionally, it accounts for potential changes in teachers' assigned classes and levels at the start of a new school year. It should be noted that if a teacher reports absent in school year 2022/2023 and remains absent in 2023/2024, the absence duration is aggregated for 2023/2024, while no new absence frequency is registered for 2023/2024. To correct inflated individual absence duration values, adjustment were made for individuals whose absence began in 2022/2023 and continued into 2023-2024. Consequently, the absence duration of 2023/2024 was subtracted, when applicable, by the latest recorded absence duration in 2022/2023. This ensures that the maximum absence duration is 365 days. Given that the Shapiro-Wilk test is highly sensitive

---

[14] CBS. (2007). *NVS Standaard voor verzuimregistratie:Nationale Verzuimstatistiek*. https://www.cbs.nl

to larger sample sizes (N >50) (Shapiro et al., 2012), normality of the endogenous variable is checked through Q-Q plots and histograms (Kolk, 2023). This study uses a significant number of independent variables. As a result, multicollinearity can occur when two or more independent variables are highly correlated, leading to complex estimations and unreliable conclusions (Henseler et al., 2024). To assess multicollinearity in regression analysis, Henseler et al. (2024) argue to use the Variance Inflation Factor (VIF), which should be smaller than ten and, even better, smaller than five.

Absence duration

The Q-Q plot and histogram (see Appendix B) indicate highly non-normal right-skewed data, which is not suitable for Ordinary Least Squares (OLS) regression (Dismuke & Lindrooth, 2006). Given the nature of HR data, non-normal distributions are common (Johns, 2008; Kamphuis, 2018; Ybema et al., 2010). In OLS regression, the assumption regarding normality applies to the residuals, not necessarily to the dependent variable itself (Rosenblatt & Shirom, 2005). To correct non-normality, the dependent variable requires transformation. Two common transformation techniques to transform a continuous (numeric) variable are: the Box-Cox and Yeo-Johnson transformation. Both techniques are frequently used in feature engineering to address skewness[15]. The Box-Cox, proposed by George Box and David Cox, applies power transformation to positive values. The output estimates the optimal lambda (λ) to achieve a normal distribution (Box & Cox, 1964). As absence duration is solely positive, it is suitable for Box-Cox transformation. The Yeo-Johnson technique is an extension of the Box-Cox method. Similarly, it applies a power tranformation and estimates the optimal λ to achieve normal distribution. Yeo-Johnson transformation can also handle negative values, whereas Box-Cox strictly positive values (Yeo & Johnson, 2000). The formulas are highlighted below.

Box-Cox

$$\psi(y,\lambda) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \lambda \neq 0, \\ \log y & \lambda = 0. \end{cases}$$

Yeo-Johnson

$$\psi(y,\lambda) = \begin{cases} \dfrac{(y+1)^\lambda - 1}{\lambda} & y \geq 0 \text{ and } \lambda \neq 0, \\ \log(y+1) & y \geq 0 \text{ and } \lambda = 0, \\ -\dfrac{(-y+1)^{2-\lambda} - 1}{2-\lambda} & y < 0 \text{ and } \lambda \neq 2, \\ -\log(-y+1) & y < 0, \lambda = 2. \end{cases}$$

[15] Tay, K. (2021). *The Box-Cox and Yeo-Johnson transformation for continuous variables*. https://statisticaloddsandends.wordpress.com/page/2/

To find the optimal λ for the Box-Cox method, a very small constant (1.0e-08) to absence duration is added since this technique only handles positive values. The optimal λ found for Box-Cox technique is ≈ 0, suggesting a normal logarithmic (log) transformation of variable Y (Box & Cox, 1964). Whereas the optimal λ for Yeo-Johnson found to be = -0.24, suggesting an inverse power function (Yeo & Johnson, 2000). Due to skewness of independent absence duration metrics and consistency, antecedents that contain absence duration are also log-transformed. These include: prior absenteeism, managerial absenteeism, and team-level absenteeism. While a log transformation is often used to transform dependent absence values (i.e. (Gellatly, 1995)), a Yeo-Johnson method can handle zero and negative values better. To analyse which method is more suitable for this research, normality of residuals (Q-Q plot), homoscedasticity, and model fit metrics are used.

First of all, the Q-Q plot of residuals (see Appendix B) for a logarithmic transformation indicate an almost normal distributed plot, with a mild deviation at the end-tail. While the Q-Q plot, using the Yeo-Johnson technique, improved the distribution compared to earlier Q-Q plots of the raw data earlier. Compared to the logarithmic Q-Q plot, it is less normally distributed. Nevertheless, according to the Central Limit Theorem, the sampling distribution of the mean becomes normally distributed as sample size is large enough (Kwak & Kim, 2017). Given the large sample size in this study (N = 1150), normality is distributed around the mean. Moreover, a Breusch-Pagan test was utilised to assess heteroscedasticity in the regression (Lyon & Tsai, 1996). The results of both tests revealed a p-value < 0.05, suggesting non-constant variance. As the OLS regression assumes homoscedasticity, no multicollinearity, and normally distributed residuals, these results violate this assumption (Dismuke & Lindrooth, 2006). To correct for this issue, robust standard errors are employed in both methods, which recognise the presence of non-constant variance (Astivia & Zumbo, 2019). All correction methods (HC0, HC1, HC2, HC3, and HC4) were used to account for heteroscedasticity (Hayes & Cai, 2007). The results remained consistent across all methods, no changes in significant predictors. Therefore, HC3 is selected as the method for applying robust standard errors due to being the most conservative.

The log-transformed model resulted in a higher adjusted R^2 compared to the Yeo-Johnson transformation (0.239 > 0.19), explaining more variation in absence duration. Moreover, the F-statistic was higher was higher (18.71 > 15.97), indicating a better model fit. Additionally, the VIF values of all antecedents in the regression are below three, indicating no multicollinearity (Henseler et al., 2024). Based on these metrics, the log transformation (using Box-Cox technique) was selected as the optimal transformation. Consequently, in the Quantitative results section, the results are based on the log transformation (using Box-Cox technique). Nevertheless, both techniques revealed exactly the same significant predictors after employing robust standard errors.

Absence frequency

To analyse the distribution of the dependent variable absence frequency, a Q-Q plot and histogram is used. The results (see Appendix B) indicate highly non-normal right-skewed data, which is not suitable for multiple linear regression (Dismuke & Lindrooth, 2006). To account for non-normality, the Poisson regression model is fitted in this study. The Poisson distribution is a Generalised Linear Model (GLM), proposed by Siméon Denis Poisson, that models the probability of events within a specific timeframe (Consul & Famoye, 2007; Coxe et al., 2008). Poisson GLM techniques do not assume normally distributed residuals and are therefore suitable for this data type. Moreover, it is the most suitable where the dependent are counts (≥ 0) (Heinzl & Mittlböck, 2003). Therefore, this technique is suitable for absence frequency since it represents the number of times a teacher reported absence in school year 2023/2024. Moreover, count variables in Ordinary Least Squares (OLS) may produce biased results when the predictor is very small (Coxe et al., 2008). The log link function in the Poisson model ensures that predicted values remain positive by applying a natural logarithmic transformation. Furthermore, it assumes that the mean and variance are equal (= 1), thus no under- or overdispersion (Coxe et al., 2008; Heinzl & Mittlböck, 2003). The dispersion test ($1.0087 \approx 1$) indicated that the mean and variance are equal, validating the choice of Poisson regression. Additionally, VIF values were all below 3, indicating no multicollinearity among predictors. To assess the model fit, R^2 cannot be computed, as it is not compatible with a Poisson regression model. Instead, Pseudo R^2 measures have been proposed to assess model fit of Poisson regression models (Heinzl & Mittlböck, 2003; Mbachu et al., 2012). While there are multiple Pseudo R^2 measures, Heinzl and Mittlböck (2003) recommend, based on an extensive simulation study, to utilise the ${R_{\mathrm{D},\gamma\mathrm{P}}}^2$ in order to measure the amount of explained variation in Poisson regression model, while accounting for dispersion. This recommendation is due to potential bias in commonly used pseudo R^2 measures. The formula is presented below.

$$R^2_{D,\gamma\mathrm{P}} = \breve{\gamma}_{\mathrm{P}} R^2_D = 1 - \frac{D(\boldsymbol{y};\mu) + k\breve{\emptyset}_{\mathrm{P}}}{D(\boldsymbol{y};\bar{y})}$$

4.1.3 Operationalisation of antecedents

This paragraph outlines the antecedents included in the quantitative analysis, highlighting the operationalisation and data transformation.

Age. The antecedent age is included, calculated by converting the date of birth into a numerical value. This approach aligns with methods used in other studies examining the role of age on absenteeism.

Gender. The factor gender is included, using a dummy variable where 1 indicates a male teacher and 0 a female teacher. This approach aligns with methods used in other studies examining the role of gender on absenteeism.

Function. Non-teaching staff individuals are excluded from the dataset. In this way, only teaching staff, are included. The existing function codes (Teacher LB, Teacher LC, and Teacher LD) in the dataset are preserved since each code represent a distinct category without any hierarchy. To handle categorical variables, the factor function in R is used.

Salary. To analyse the role of salary on teacher absenteeism, salary scales are included. These scales include Teacher LB, Teacher LC, Teacher LD, while salary step data was not available[16]. As salary levels represents an ordinal relationship, they are treated as ordinal variables. However, ordinal scales assume equal spacing between salary levels. To address this, salary levels are encoded to preserve ordinal relationship without assuming equal spacing. The assigned codes are as follows: (1) LB, (2) LC, and (3) LD.

Working hours (FTE). The variable working hours (full-time equivalent) represent the proportion of a full-time workload assigned to each teacher, ranging from 0 to 1. Where 1 indicates a full-time position and values below 1 represent a part-time employment. The dataset contains average FTE, representing the weighted average of the FTE over the selected period excluding a reduction in working hours for older secondary education staff (BAPO). As FTE is a continuous variable, it can be directly included in the analysis without transformation.

FTE in team. FTE in team represent the aggregated FTE for each team. In the dataset, some teachers are assigned to multiple teams leading to inconsistencies in calculating the FTE in team. Based on amount of FTE spend in each team, data is normalised. First of all, rows where a teacher contributes 0 FTE to a team are removed. The remaining FTE values were

[16] Rijksoverheid. (2024). *Scales, steps and rewards in secondary education*. https://www.rijksoverheid.nl/onderwerpen/werken-in-het-onderwijs/vraag-en-antwoord/wat-verdien-ik-als-leraar-in-het-voortgezet-onderwijs

normalised to reflect the proportion of a teachers' FTE spent in each team. This normalisation ensures that the FTE values across all teams for each individual sums to 1. For example, if a teacher (0,67 FTE in total) spends 0,32 FTE in team A and 0,35 FTE in team B, the normalised values are: Team A: 0,32 / 0,67 = 0,477 and Team B: 0,35 / 0,67 = 0,522.
Using the normalised values, the weighted FTE in each team is calculated, by multiplying the normalised FTE by the total FTE of the corresponding team. Thus: Team A: 0,477 * 16,47 (total FTE team A) = 7,866 and Team B: 0,477 * 16,81 (total FTE team B) = 8,781. The final FTE in team for this example is the sum of these weighted values 16,65 FTE (7,866 + 8,781). The normalisation and weighting addresses redundancy and incorrect calculations caused by multiple rows. Thus, ensuring that each individual is represented by a single FTE in team value, suitable for further analysis.
Employment status. This antecedent is currently stored as string data, indicating whether an employee works full time (true) or part-time (false) based on their FTE value. For statistical analysis, this antecedent is recorded into a dummy variable, where (1) indicated full-time employment and (0) part-time employment.
Team level absence. Team level absence represent the aggregated absence duration and frequency within each team. In the data, some individuals are assigned to multiple teams, which can lead to inconsistencies in calculating team-level absence metrics. Similarly as for FTE in team, data is normalised based on the proportion of time each individual spends to each team. First of all, rows where a teachers contributes 0 FTE to a team are removed. The normalisation ensures that FTE values across all teams for each individuals sums to 1 and are distributed proportionally to their contribution. To remain consistent throughout this study, maternity and planned leave are also excluded for the total amount of team metrics. Additionally, the total amount team-level absence metrics includes duration and frequency data from non-teaching staff. Although these individuals are not included in the primary analysis, they remain part of the total team-level metrics since they can be members of the same team. Thus, ensuring that the overall metrics remain complete. For example, if a teacher (0,8 FTE in total) spends 0,1 FTE in team A and 0,7 FTE in team B, the normalised values are: Team A: 0,1 / 0,8 = 0,125 and Team B: 0,7 / 0,8 = 0,875. The weighted absence duration for this individual is calculated by multiplying the normalised values by the total team-level absence. Thus, Team A: 0,125 * 2304 (absence duration team A) = 288 and Team B: 0,875 * 2030 = 1776,25. The final team-level absence duration for this individual is the sum of these weighted values 2064,25 (288 + 1776,25). Similarly, the team-level frequency is calculated, Team A: 0,125 * 47 (absence frequency team A) = 5,875 and Team

B: 0,875 * 55 = 48,125. Consequently, the final team-level absence frequency for this individual is the sum of these weighted values 54 (5,875 + 48,125). This approach addresses redundancy and incorrect calculations caused by multiple rows, ensuring that each individual is represented by a single adjusted team-level absence value, suitable for further analysis.

Multiple schools. This antecedent, derived from contract data, indicates whether teachers are employed at multiple schools. Individuals with 0 FTE at a school were excluded from the analysis. The antecedent was subsequently recorded into a binary measure, where teachers working at multiple school are codes as (1), and those with one school as (0).

Menopause. According to World Health Organisation (WHO) menopause usually occurs between the age range 45-55 years old[17]. Although, research indicates that not all women experience menopausal symptoms or experience it earlier or later (O'Neill et al., 2023; Verdonk et al., 2022). Due to limited availability of specific menopausal data, this study adopts a general proxy based on typical age ranges and gender. Therefore, this approach may include women without symptoms and exclude women outside the age range. For quantitative analysis, a dummy variable is created, where (1) indicates menopause for teachers identified as women within the age range of 45-55 years, and (0) all others.

Prior absenteeism. This variable captures absenteeism from the previous school year (2022/2023) in relation with the timeframe of this study (school year 2023/2024). Although, situations may arise where teachers left the school after 2023 or joined in 2024. This can introduce missing data or bias in the analysis, and are therefore excluded from the analysis.

Managerial Absence. Manager absence represents the individual absence values of managers in relation to each individual teacher. As mentioned, this study is analysed from two perspectives, the total absence duration and the amount of absence frequencies. Therefore, this variable is measured as the sum of all recorded absence days for a manager in this study timeframe. Similarly, absence frequency represents the total number of absence events recorded. Since these variables are continuous, it can directly be included in the analysis.

Team level absence. Team level absence represents the aggregated absence duration for each team, calculated by summing the individual absence values of all members in a team. Similarly as for managerial absence, two perspectives are used, the total absence duration and the amount of absence frequencies. Team level absence is also aggregated by the absence

[17] WHO. (2024). *Menopause*. https://www.who.int/news-room/fact-sheets/detail/menopause#:~:text=Most%20women%20experience%20menopause%20between,in%20circulating%20blood%20oestrogen%20levels.

frequency for each team, calculated by summing the individual absence frequencies of all team members in a team. Since both values are continuous, it can be directly included in the analysis without transformation.

Age of manager. The antecedent age of manager represents the age of the manager assigned to each employee. This is calculated by converting the manager's date of birth into a numerical value. Although prior research has not directly addressed the relationship between a manager's age and absenteeism, a data-driven approach is used to include this variable. Given it's a continuous value, it can be directly incorporated in the analysis.

Tenure. Tenure represent the amount of time an employee has been employed within the organisation in years. To calculate tenure, the data column was transformed in a proper date format. Given that employees had multiple employment contracts, only the oldest start date was retained. Consequently tenure was calculated by the time difference between the oldest contract start date and the end of school year 2023/2024. Tenure is represented in decimal years, by computing the total number of months employed and converting it into years (e.g., 5.25 years for 5 years and 3 months).

Contract. This antecedent represent the contract type of a teacher, distinguishing fixed and permanent contracts. Some individuals had multiple contracts, including both fixed and more recent permanent contracts. However, this is due to additional minor work while the majority of their FTE was associated with a fixed contract. To ensure consistency, these individuals were classified under the fixed contract category which better reflects their employment status. Moreover, in cases where individuals were employed at multiple schools with both fixed and permanent contracts, fixed contract types were prioritised. Suggesting that one fixed contract type offers a more stable employment status. For statistical analysis, this antecedent is recoded into a dummy variable. Where (1) indicates fixed contracts and (0) permanent contracts.

Municipality. This antecedent represents the school location in which the teacher work. The school location is then recoded into a dummy variable, where (1) indicates towns and (0) villages. The classification based on school location has been pre-specified by the educational institution from which the data is utilised. In the dataset, 17 individuals (≈1.4% of sample size) worked at multiple schools, one school in a town and one in a village. As a result, a 'mixed' category was avoided due to potential statistical issues and the risk of overfitting. Instead, an FTE-weighted approach was applied. For individuals working at multiple schools, the classification is determined by the location where the teacher spends the majority of their FTE. There was one individuals who spend the same amount of FTE at two schools.

Consequently, this individual was assigned to the school with the first recoded contract-date, providing a more objective tie-breaker without randomness.

Replacement/Temporary vacancy. This variable, based on employment reason data, captures whether a teacher's employment includes fulfilling a temporary vacancy, serving as a replacement, or both. For statistical analysis, this antecedent is recorded into a dummy variable. Where (1) indicates replacement/temporary vacancy and (0) individuals with no such contracts. This classification includes where these contracts are either the individual's only employment or exist alongside their primary contract. This approach takes into account the potential employment instability with temporary vacancies or replacement roles.

School type. This antecedents represents different school types based on school data. Similarly as for the municipality antecedent, an FTE-weighted approach was applied, given that some individuals work at multiple schools. For individuals working at multiple schools, the school is determined where the teacher spends the majority of their FTE. There was one individuals who spend the same amount of FTE at two schools. Consequently, this individual was assigned to the school with the first recoded contract-date, providing a more objective tie-breaker without randomness. In the Netherlands, secondary education consists out of four different levels, excluding specialised education: practical education (pro), pre-vocational secondary education (vmbo), senior general secondary education (havo), and pre-university education (vwo)[18].

The school types have been pre-specified by the educational institution from which the data is utilised. This consists out of the four distinct school types:

- Multi-level schools (vmbo/havo/vwo): schools that offer multiple educational levels
- Track-specific schools (havo/vwo or vmbo/havo): schools that provide limited selection of education tracks
- Vmbo schools: schools that only provides vmbo level
- Practical education: schools that only provides practical education

Control antecedents

To account for potential bias in the analysis, control antecedents are incorporated to minimise confounding effects. Studies have consistently included age and gender as control

[18] Ministry of education, culture, and science. (2024). *Schematic representation of the Dutch education system*. https://www.ocwincijfers.nl/sectoren/onderwijs-algemeen/schooltypen/schooltypen#:~:text=Het%20voortgezet%20onderwijs%20bestaat%20uit,voorbereidend%20wetenschappelijk%20onderwijs%20(vwo).

antecedents to account for demographic differences in absenteeism patterns (Kamphuis, 2018; Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991). Highlighting that female teachers show higher absenteeism rates, possible due to family responsibilities or health-related factors. Additionally, studies consistently included working hours (FTE) as a control to minimise confounding effects. Teachers with higher FTE may experience higher absenteeism rates, have additional responsibilities, encounter more early starts and accumulate gaps since they have more total hours scheduled (Barmby et al., 2001; Possenriede et al., 2014). Accordingly, these variables are controlled.

4.2 RESULTS

This study is conducted from two dependent variables, total absence duration and absence frequency. Consequently, the results are described along these variables. As this study explores new antecedents that have not been explored by existing literture, a p-value < 0.1 was considered a trend (marginally significant). This approach minimises Type II errors (failing to detect a true effect) that meet conventional thresholds ($p < 0.05$).

To assess group differences in antecedents such as teacher function and school type, appropriate statistical tests were selected based on the distribution and variance characteristics of the data. A Levene's test was conducted to test for equal variances since it does not assume normality. For both absence duration and frequency, both tests found to be significant ($p < 0.05$), suggesting unequal variances. Consequently, to test differences between groups (function and school type) a regular analysis of variances (ANOVA) cannot be conducted while it assumes equal variances. Since the data itself is non-normal, Hangcheng (2015) recommends to conducted a non-parametric ANOVA test, more specifically the Kruskal-Wallis test, as it does not assume equal variances and normality of the dependent variable. Consequently, if statisically significant effects were observed, post-hoc tests were conducted to test which groups differ from each other. Dunn's post-hoc test is commonly used after the Kruskal-Wallist test (Dunn, 1964). However, this test does not take into account control variables. Assuming normally distributed residuals, a pairwise comparison that accounts for covariates included in the regression model was conducted. More specifically, an Estimated Marginal Means (EMMs) post-hoc comparison (Lenth, 2025). The p-values are adjusted using the Tukey-Kramer technique, proposed by Tukey (1977), given that sample sizes are unequal. This post-hoc test is suitable for both OLS regression and GLM Poisson regression (Lenth, 2025). While regression models primarily test differences between each category and a single reference category, the EMMs post-hoc offers a more comprehensive analysis of category differences by comparing all types against each other. Thus, results are not biased by the selection of the reference category and Type I error is controlled (Lenth, 2025). Therefore, if significant predictors are found, the post-hoc results are presented in the regression table. Moreover, while standardised coefficients allow for comparison of relative predictor strength, unstandardized coefficients are reported in this study to retain interpretability in real-world terms. Table 7 presents the significant antecedents of teacher absenteeism found in this study.

Table 7 Regression Results

(unstandardised coefficients)	**Absence Duration**	**Absence Frequency**
Personal		
Age	0.008*	
Gender (Male)	-0.318**	-0.161**
Gender × Age		
Menopause		
Prior absenteeism	0.435***	0.173***
Job-related		
Working hours (FTE)	1.232***	1.400***
Gender × Working hours		
FTE in team	-0.024***	-0.042***
Function		
Gender × Function		
Multiple schools		
Gender × Multiple schools (Male)	1.013 .	
Age of manager		-0.009*
Managerial absence		
Tenure	-0.026***	-0.015**
Team-level absence	0.094 .	0.025***
Municipality		
Replacement/Temporary vacancy		
School type (track-specific vs. vmbo)	-0.386**	
(multi-level vs. vmbo)		0.231 .
Socio-economic		
Salary scale		
Contract type (Fixed)		0.214***
Employment status (Full-time)	-0.585***	-0.317***
Adj. R-Squared	0.239	0.218

Note: ‘***’ $p < 0.001$, ‘**’ $p < 0.01$, ‘*’ $p < 0.05$, ‘.’ $p < 0.1$ (marginally significant). An empty cell for the antecedent indicates that no significant relationship was found.

Table 8 provides an overview of the hypotheses tested in this study, along with the observed relationships for each antecedent for absence duration and absence frequency. This summary allows for a quick comparison between the hypothesised relationship and the results, highlighting with antecedents align with the expectations and which deviate, thus refining our understanding of teacher absenteeism.

Table 8 Hypotheses Overview and Results

	Antecedent	Hypotheses		Absence duration	Absence frequency
Personal	Age	H1	-	+	
	Gender *(female)*	H2	+	+	+
	Prior absenteeism	H3	+	+	+
	Menopause	H4	+		
Job-related	Tenure	H5	+	-	-
	Working hours (FTE)	H6	+	+	+
	Function	H7	+/-		
	Age of manager	H8	+/-		-
	Managerial absence	H9	+		
	Multiple schools	H10	+	+ *(males only)*	
	FTE in team	H11	+/-	-	-
	Team-level absence	H12	+	+	+
	Municipality	H13	+/-		
	Replacement/temporary vacancies	H14	+/-		
	School type:	H15			
	- *track-specific vs. vmbo schools*		+/-	-	
	- *multi-level vs. vmbo schools*		+/-		+
Socio-economic	Salary	H16	-		
	Contract type *(permanent)*	H17	-		-
	Employment status *(full-time)*	H18	-	-	-

Note: '+' indicates a positive relationship, '-' a negative relationship, and '+/-' means a non-directional relationship, implying a difference is expected though the direction of the relationship remains unclear. An empty cell in the column 'Absence duration' or 'Absence Frequency indicates that no significant relationship was found

4.2.1 Absence duration

The OLS regression identified 10 significant variables as antecedents of absence duration. Given that absence duration has been log-transformed, the β-coefficients are exponentiated to interpret percentage changes. The first variable *age* tested a positive significant relationship ($\beta = 0.008$, $p < 0.05$). Contrary to the hypothesis (H1), older teachers have higher absence durations compared to younger teachers. Each additional year of age is associated with 0.88% longer absence duration, holding all other variables constant. Second, *gender* has a statistically significant effect on absence duration, revealing that male teachers have shorter absence duration (24,5%) than female teachers ($\beta = -0.274$, $p < 0.001$) for those working at a single school. This result is in line with the hypothesis (H2). Teachers with high *absence* *duration in the previous school year* also exhibit significantly longer absences in the next school year ($\beta = 0.435$, $p < 0.001$). A 1% increase in last school year's absence duration is associated with a 0.43% increase in this year's absence duration, holding all variables constant. This conforms the initial expectation (H3). Moreover, multiple job-related factors play a crucial role in absence duration. Contrary to hypothesis (H5), *tenure* is negatively related to absence duration ($\beta = -0.026$, $p < 0.001$), highlighting that teachers with longer tenure have shorter absence durations. The amount of working hours (*FTE)* is positively and significant associated with absence duration ($\beta = 1.232$, $p < 0.001$), indicating that employees with more working hours tend to have longer absence durations, aligning with the hypothesis (H6). Conversely, the proportion of *FTEs in a team* is negatively associated with absence duration ($\beta = -0.024$, $p < 0.001$), demonstrating that teachers working in larger teams experience shorter absences. This conform the hypothesis that a relationship will be found (H11). A marginally significant relationship was found between *team-level* absence duration and individual absence duration ($\beta = 0.094$, $p < 0.1$), confirming the expectation (H12). Proposing that higher team-level absenteeism may contribute to increased individual absenteeism, potentially adjusting their. personal absence norm with the group's absence norm. The socio-economic factor employment status (full-time vs. part-time) is also a significant determinant of absence duration ($\beta = -0.585$, $p < 0.001$). Full-time teachers have shorter absence durations compared to part-time teachers, confirming the hypothesis (H18).

The Kurskal-Wallis test showed a significant difference in absence duration across school types ($\chi^2 = 20.747$, $p = 0.0001$). The EMMs post-hoc comparison, based on the OLS regression model, using Tukey-Kramer adjustment revealed significant differences in absence duration between track-specific schools and vmbo schools ($\beta = -0.386$, $p = 0.038$), while controlling for covariates. More specifically, track-specific schools had significantly lower absenteeism compared to vmbo schools. No significant differences were found between other school types and absence duration. Moreover, the OLS regression also revealed a significant difference between track-specific schools and vmbo schools, with track-specific schools as the reference group. This confirms the expectation that a relationship will be found (H15). The Kurskal-Wallis test showed no significant difference between function (Teacher LB, Teacher LC, and Teacher LD) and absence duration ($\chi^2 = 2.077$, p-value = 0.354). Moreover, the regression model reveals no significance, indicating no difference between absence duration and function. Thus, rejecting the hypothesis (H7) that a relationship will be found.

Prior absenteeism literature (Brooke & Price, 1989; Ost & Schiman, 2017; Rosenblatt & Shirom, 2005) have tested interaction effects (such as, gender × age, gender × function , gender × FTE, team-level × managerial absence), but were found to be non-significant in this study and therefore are excluded from the OLS model (all $p > 0.1$). Given this study includes new antecedents that have not been explored by current literature, interaction effects have been explored without theoretical justification. A marginally significant interaction between working at multiple schools and gender was found ($\beta = 1.013$, p-value = 0.069), therefore partially accepting the hypothesis (H10). The main effect of working at multiple schools was not significant, meaning working at multiple schools is not significantly associated with absence duration. The interaction effect highlights that male teachers who work in multiple school have significantly higher absence duration compared to male teachers in single schools. Absence duration for female teachers is not significantly influenced by working in multiple schools. However, the 95% confidence interval is wide (0.08, 1.80), indicating a high degree of uncertainty. Given the small number of teachers working in multiple schools ($n = 43$), this finding should be interpreted with caution and validated in future studies with larger samples. After introducing this interaction in the OLS regression, the overall improvement in model fit was not significantly improved Nevertheless, a slightly higher adjusted R^2 was observed and the interaction effects remained in the model due to its practical implication. No significant relationship with absence duration was found for hypotheses H4, H8, H9, H11, H13, H14, H16, and H17. Therefore, these hypotheses are rejected in this study for absence duration.

The adjusted R^2 (0.239) of the OLS regression model suggests a good model fit, where 24% of the variation in the dependent variable is explained by the antecedents. Moreover, the F-statistic (18.71) and p-value for the F-test ($p < 0.05$) indicate strong model significance.

4.2.2 Absence frequency

The Poisson regression identified ten significant variables as antecedents of absence frequency. Poisson model coefficients cannot be interpreted as linear interpretation, and therefore should be exponentiated giving Incidence Rate Ratios (IRRs) (Coxe et al., 2008). The first significant variable, *gender*, reveals that male teachers have significant lower absence frequencies than female teachers ($\beta = -0.161$, $p < 0.01$). Exponentiating this implies that male teachers have 14.9% fewer absence frequencies than female teacher, holding other factors constant. This is consistent with the hypothesis (H2). Second, *prior absence frequency* is significantly associated with absence frequency ($\beta = 0.173$, $p < 0.001$), revealing that teachers who had high absence frequenies in the previous school year are more likely to take frequent absences in the current school year. This conforms to the initial expectation (H3). Contrary to the hypothesis (H5), *tenure* was found to be negatively related to abence frequency ($\beta = -0.015$, $p < 0.01$), highlighting that for every additional year of tenure, the expected frequency of absences decreases by 1.5%. Moreover, working hours (*FTE)* found to be positively and significant associated with absence frequency ($\beta = 1.400$, $p < 0.001$), meaning that teachers with more working hours have higher absence frequencies, supporting the anticipated outcome (H6). Interestingly, the managerial characteristic age of manager plays a critical role. The *age of mananger* is negatively associated with absence frequency ($\beta = -0.009$, $p < 0.05$), suggesting that teachers working under older managers have slightly lower absence frequencies. Confirming the hypothesis (H8) that a relationship will be found. In addition, the proportion of *FTEs in team* is negatively and significant related to absence frequency ($\beta = -0.042$, $p < 0.001$). Implying that for each additional FTE in team, an teacher's expected absence frequency decreases by approximately 4.1%. This aligns with the hypothesis (H11) that a relationship will be found. The seventh significant predictor *team-level absence frequencies* is postively related to individual absence frequency ($\beta = 0.025$, $p < 0.001$). Demonstrating that higher absence frequencies in teams contributed to higher individual absence frequencies, confirming the hypothesis (H12). Teachers potentially adjust their personal absence norm with the group's absence norm. The first socio-economic antecedent, *contract type,* is a significant determinant of absence frequency ($\beta = 0.214$, $p < 0.001$). Teachers with fixed-term contracts have significantly higher absence frequencies (23.8%) than those with permanent contracts, aliging with the expectation (H17). Additionally, *employment status* (full-time vs. part-time) is a strong predictor of absence frequency ($\beta = -0.317$, $p < 0.001$). Full-time teachers take significantly fewer absence frequencies (27.1%) than part-time teachers. This is also consistent with the hypothesis (H18).

The Kurskal-Wallis test showed a significant difference in absence frequency across school types ($\chi^2 = 11.735$, p-value = 0.008). The Poisson regression found significant differences between multi-level schools and vmbo schools ($p < 0.05$). These results are only compared from the reference category multi-level school, as that is the largest sample size. In contrast, the post-hoc pairwise comparison using EMMs with Tukey-Kramer adjustment, while

controlling for covariates, found a marginally significant difference between multi-level schools and vmbo schools at the 0.10 level ($\beta = 0.231$, $p = 0.086$). More specifically, vmbo schools had marginally significant lower absence frequency compared to multi-level schools. This outcome support H15, which proposed a significant association. No significant differences were observed between other school types and absence frequency.
Similarly as for absence duration, The Kurskal-Wallis test showed no significant difference between function (Teacher LB, Teacher LC, and Teacher LD) and absence frequency ($\chi^2 = 4.459$, p-value = 0.108). Moreover, the regression model reveals no significance, indicating no difference between absence frequency and function. Indicating no difference between absence frequency and function, thus rejecting the hypothesis (H7). In addition, no significant relationship with absence duration was found for hypotheses H1, H4, H9, H10, H13, H14, and H16. Therefore, these hypotheses are rejected in this study for absence frequency.

Prior absenteeism literature ((Brooke & Price, 1989; Ost & Schiman, 2017; Rosenblatt & Shirom, 2005)) have tested interaction effects (such as gender × age, gender × function , gender × FTE), but were found to be non-significant in this study and therefore are excluded from the final model (all $p > 0.05$). Only gender × age was found to be significant for absence frequency. However, the VIF values was greater than 18, indicating multicollinearity. According to Robinson and Schumacker (2009) when examining interaction effects, centering of variables is required since uncertain variables impact the VIF values. After centering age, the VIF value dropped below 3, however the interaction effect became non-significant and therefore was also removed from the analysis. The final model only includes statistically significant predictors, ensuring optimisation and explanatory power.

The $R_{D,\gamma P}{}^2$ value (0.218) in the poisson regression on absence frequency, suggests a good model fit, 22% of the variation in the dependent variable is explained by the antecedents. Moreover, a substantial reduction in deviance was found compared to the null model (399.2), indicating a strong model significance.

5. STUDY 3 - MACHINE LEARNING PREDICTION

Building on the significant antecedents identified in Study 2, Study 3 (highlighted in blue) employs Machine Learning (ML) models to predict individual teacher absenteeism categories. The aim is to assess the predictive accuracy using validated antecedents. Figure 8 presents the overall multi-study approach adopted in this research.

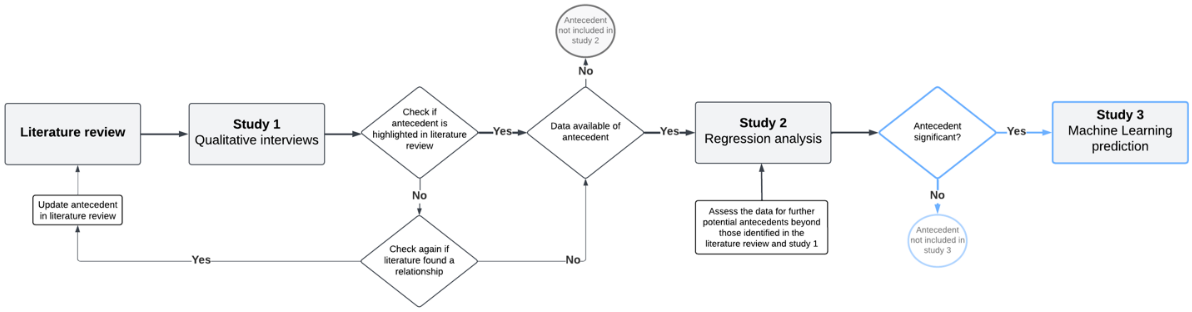

Figure 8 Overview of the Multi-Study Research Approach (Study 3)

While Study 2 highlights which antecedents are statistically significant predictors of absenteeism, Study 3 offers additional practical value by focusing on how well these predictors can classify individual teachers into absence categories. This aligns with the growing practical demand of predictive absenteeism analytics in the educational sector (Hoppenbrouwer & Bouma, 2021). Although previous studies have ML algorithms to predict absenteeism (Ajmi, 2019; Fernandes & Filho, 2021; Gayathri, 2018; Montano et al., 2020), their application is lacking within the educational context. In this study, Naïve Bayes and Support Vector Machine (SVM) algorithms are applied to predict teachers' absence duration categories based on the same dataset as Study 2. The models use validated antecedents from Study 2 as model features (see Table 9). This prediction enables school administrators to proactively identify which teachers are most at risk and tailor individual interventions more effectively to potentially reduce overall absenteeism. The following sections detail the methodological approach for training and evaluating these ML models, as well as the resulting performance metrics.

Table 9 Model Features

Model Features		
Personal	***Job-related***	***Socio-economic***
Age	Working hours	Contract type
Gender	FTE in team	Employment status
Prior absenteeism	Multiple schools	
	Age of manager	
	Tenure	
	Team-level absence	
	School type	

5.1 METHOD

While the regression results provides insights into the significant predictors of absenteeism, a more in-depth predictive analysis is essential to enhance the practical contribution. Therefore, this study applies ML algorithms to classify absence duration. The classification is based on the Dutch NVS (absenteeism statistic)[19] framework. They classify the following three classes: short-term (≤ 7 days), medium (≥ 8 and ≤ 42 days), and long-term (≥ 43 and ≤ 365 days).

Previous research have examined and applied Machine Learning (ML) algorithms to predict absenteeism (Ajmi, 2019; Gayathri, 2018; Montano et al., 2020). The most commonly used algorithms for predicting absenteeism include Support Vector Machine (SVM), which determine the best boundary to separate classifications in the data. Naïve Bayes, based on the Bayes' Theorem, predicts categories by assuming feature independence. And supervised Neural Network (NNs), that consists of layers of neurons that learn complex patterns like human brains (Mahesh, 2020) These algorithms are supervised ML techniques, that map an input (absenteeism) to an output based on example input-output pairs (Mahesh, 2020).

A rule of thumb for NNs is that the minimum sample size should be fifty times the number of weights (Alwosheel et al., 2018). It can be calculated using the following formula:

Total Weights = (Input neurons + 1) x Hidden neurons + (Hidden neurons + 1) x Output neurons

Study 2 identified twelve significant predictors from the regression result that serve as input neurons. The output neurons contains three categories (short-term, medium, and long-term) and assuming one hidden layer with 10 neurons, the total number of weights equals 160. Applying the fifty-time rule of thumb, the total dataset for reliable NN training should be at least 8.000 samples. Given the dataset size of this study (N = 1150), this ML algorithms has not been applied. Consequently, this study applies Naïve Bayes and SVM algorithms to classify absenteeism categories based on the significant predictors from the regression result. These models are more suitable for smaller datasets (Ajmi, 2019; Montano et al., 2020). This study applies a 70% training and 30% test data split, as this is commonly used in ML.

---

[19] CBS. (2007). *NVS Standaard voor verzuimregistratie:Nationale Verzuimstatistiek*. https://www.cbs.nl

To evaluate the models, the following key performance metrics are used, accuracy, F1 score, ROC (Receiver Operating Characteristic) curve, AUC (Area Under the Curve) value, and the confusion matrix (Varoquaux & Colliot, 2023). F1 score provides a balance between precision and recall, while the confusion matrix provides visual insights into the true positives (TP) , false positives (FP), true negatives (TN), and false negatives (FN). The ROC curve assesses model performance across different thresholds, while the AUC value quantifies the ability to distinguish classes, higher values mean better performance (Ajmi, 2019; Varoquaux & Colliot, 2023). Given that R studio is used in Study 2, it is also utilised for training, testing and evaluating the ML models.

5.2 RESULTS

Given that this study applied SVM and Naïve bayes algorithms to classify NVS absenteeism categories based on absence duration (short-term (≤ 7 days), medium (≥ 8 and ≤ 42 days), and long-term (≥ 43 and ≤ 365 days). The results are described along these algorithms. The confusion table and key performance metrics for both algorithms are presented in Table 10.

Table 10 Confusion matrices and Performance metrics

Naïve Bayes					**Support Vector Machine (SVM)**				
		Actual					***Actual***		
		Short-term	Medium	Long-term			Short-term	Medium	Long-term
Prediction	Short-term	224	4	11	***Prediction***	Short-term	222	57	0
	Medium	7	66	0		Medium	14	17	3
	Long-term	5	5	22		Long-term	0	1	30
Accuracy		90.7%			**Accuracy**		78.2%		
F1 Score		0.94			**F1 Score**		0.86		
Average AUC		0.97			**Average AUC**		0.88		

Based on the results, the Naïve Bayes model provided better results compared to the SVM model. The Naïve Bayes model achieves 90.7% accuracy and F1 score of 0.94, indicating excellent model performance. Moreover, the ROC curve (see Appendix C) shows a steep curve that hugs the top-left corner, confirmed by the average AUC (Area Under Curve) value. Indicating that the classifier accurately (0.97) distinguishes between absenteeism categories.

As shown in the confusion matrix, short-term absences (224 correct, 12 misclassifications) and medium (66 correct, 9 misclassifications) are identified correctly in most cases. Long-term absences are identified noticeably lower than for the other classes. Potentially, due to imbalance in class distribution compared to short-term and medium absence categories. In contrast, the SVM algorithm provided significantly lower accuracy (78.2%), F1 score (0.86) and average AUC (0.88). Additionally, the ROC curve (see Appendix C) is especially lower for short-term and medium absence categories. As shown in the confusion matrix, while it correctly identified a large proportion of short-term and long-term absences, it misclassified 57 medium cases as short-term. These results shows that algorithm choice matters, particularly in smaller datasets. Therefore, the Naïve Bayes algorithm is preferred for real-world application in the educational sector to classify absence duration categories.

These metrics indicate that the ML algorithms can provide actionable foresights into which teachers are likely to fall into longer absenteeism categories. Unlike traditional statistical model (Study 2) that only indicates associations, this study adds value by enabling individual-level predictions based on contextually validated antecedents. This enables schools to anticipate potential absenteeism patterns and intervene earlier. The model can be integrated into existing HR dashboards to move from descriptive to predictive analytics, a shift that shows growing demand (Hoppenbrouwer & Bouma, 2021), especially during teacher shortages [20].

[20] Dutch Ministry of Education, Culture, and Science (2023). *Trendrapportage Arbeidsmarkt Leraren po, vo en mbo 2023*. https://www.aob.nl/assets/Bijlage-3.1-Trendrapportage-Arbeidsmarkt-Leraren-2023.pdf

6. DISCUSSION

This study examined the antecedents of teacher absenteeism in Dutch secondary education using a multi-method sequential research design that integrated qualitative interviews (Study 1), regression analysis (Study 2), and ML algorithms (Study 3). This approach ensured to capture both subjective experiences and statistical associations. Table 11 presents a comprehensive overview of antecedents influencing teacher absenteeism as identified across the literature, Study 1, and Study 2. The overall findings of this research provide a better understanding of how personal, job-related, and socio-economic antecedents shape teacher absenteeism, highlighting both consistency and discrepancies in their significance and direction. Furthermore, the studies contributed by identifying previously unexplored antecedents and providing predictive modelling to support HR decision-making. It is important to note that several antecedents, such as health status, stress, job satisfaction, and having children, although confirmed in Study 1 and supported by literature, were not included in Study 2 due to data limitations. This restricts the ability to statistically validate these predictors and highlights a key area for future research using more comprehensive datasets. Moreover, due to the multi-faceted nature of absenteeism, there might be antecedents that remain unexplored in this research.

Table 11 Antecedents across Studies

	Antecedent	**Literature**	**Study 1**	**Study 2**		**Study 2 - Result**	
				Hypotheses		*Duration*	*Frequency*
Personal	Age	-	-	H1	-	+	x
	Gender *(female)*	+	+	H2	+	+	+
	Marital status *(married)*	+					
	Education	-					
	Prior absenteeism	+		H3	+	+	+
	Children	+	+				
	Health-status	-	-				
	Menopause	+	+	H4	+	x	x
	Perfectionism		+				
	Busy social life		+				
Job-related	Tenure	+		H5	+	-	-
	Working hours (FTE)	+	+	H6	+	+	+
	Function			H7	+/-	x	x
	Experience	+	-				
	Job-satisfaction	-	-				
	Working environment/conditions	-	-				
	Job involvement	-					

Table 11 Cont.

	Antecedent	**Literature**	**Study 1**	**Study 2**		**Study 2 - Result**	
				Hypotheses		*Duration*	*Frequency*
Job-related	Autonomy	-	-				
	Commuting distance	+					
	Stress	+	+				
	Leadership style *(supportive)*	-	-				
	Additional responsibilities	+	+				
	Monitoring/Inspection	-	+				
	Age of manager			H8	+/-	x	-
	Managerial absence	+		H9	+	x	x
	Multiple schools		+	H10	+	+ *(males only)*	x
	FTE in team			H11	+/-	-	-
	Team-level absence	+	+	H12	+	+	+
	Student behaviour	+	+				
	Digitalisation		+				
	Flexibility		-				
	Following a study		+				
	Less favourable schedule		+				
	Hierarchical organisation		+				
	Subject		+/-				
	Municipality		+/-	H13	+/-	x	x
	Replacement/temporary vacancies			H14	+/-	x	x
	School type:			H15	+/-		
	- *track-specific vs. vmbo schools*					-	
	- *multi-level vs. vmbo schools*						+
Socio-economic	Salary	-		H16	-	x	x
	Contract type *(permanent)*	-		H17	-	x	-
	Employment status *(full-time)*	-		H18	-	-	-
	Tight labour market		+				
	Tight housing market		+				
	Parental influence		+				
	Social safety		-				
	Public opinion		+				

Note: '+' indicates a positive relationship, '-' a negative relationship, and '+/-' means a non-directional relationship, implying a difference is expected though the direction of the relationship remains unclear. An 'x' in the column 'Study 2 - Results' means that no significant relationship was found, an empty column in this cell means that no data was available to test the relationship.

In the subsequent sections, we critically synthesise the results across studies in relation to existing literature, outline the research contribution to both academia and professionals, acknowledge limitations, and propose future research directions. The chapter concludes by presenting a summary that answers the research questions.

6. 1 SUMMARY OF FINDINGS AND IMPLICATIONS

The following synthesises the results across each study in relation to the literature to identify consistencies, discrepancies, and unique insights into teacher absenteeism.

6.1.1 Triangulated Antecedents: Literature, Study 1, and Study 2

Our findings confirmed several antecedents of teacher absenteeism that were consistently supported across literature, Study 1 and Study 2. Highlighting that these antecedents also hold true in Dutch secondary education, thereby supporting and reinforcing existing empirical evidence in absenteeism literature. Consequently, support Hypotheses 2, 6, and 12. The antecedents *gender*, *working hours*, and *team-level absence* emerged as robust predictors of teacher absenteeism. First of all, female teachers were found to have significantly higher absenteeism, consistent across both absence duration and frequency. These results are consistent with previous research (Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991), highlighting the key role of *gender* in absenteeism. Secondly, our findings confirmed that more contracted *working hours* are associated with increased absenteeism, as identified by Rosenblatt and Shirom (2005) and Jacob (2010). Implying that educational institutions need to carefully manage teacher workloads to reduce absenteeism and promote sustainable employability. Although, Virtanen et al. (2011) found that women moderate this effect, Study 2 did not identify a moderation effect. Moreover, our findings consistently confirmed *team-level absenteeism* as a key antecedent of teacher absenteeism with both absence measures, supporting prior research from Geurts et al. (1994) and Steers and Rhodes (1978). Study 1 participants described that a low reporting threshold and normalised absence behaviour, particularly within in smaller teams, affect absenteeism. Implying that teacher tend to adjust their personal absence norm with the group's absence norm, supporting the concept of an absence culture. These three antecedents shows robust relationships with teacher absenteeism, validated through theory, experience, and statistical analysis.

6.1.2 Theory backed evidence: Literature and Study 2

The following antecedents are in line with the hypotheses, but are not identified in Study 1. Given that the findings from Study 2 are grounded by quantitative analysis and aligned with existing literature, they provide a validated conclusion. Nevertheless, the aim of Study 1 was to identify antecedents not mentioned by previous literature. Consequently, we find support for Hypotheses 3, 17, and 18.

First of all, *prior absenteeism* was found to be a significant predictor for both duration and frequency. These results are in line with the results obtained by Keller (1983), and Rosenblatt and Shirom (2005). Implying that absenteeism in the previous school year highly affects future absenteeism. While this antecedent was not mentioned in Study 1, potentially due to participant not recognising prior absenteeism as a contributing factor. Moreover, this research shows the relationship between *contract type* and absence frequency, addressing that teachers with temporary contracts have significantly less absenteeism compared to teachers with fixed contracts. These findings are in line with those from Bradley et al. (2007). Implying that contract teachers benefit from greater job security and experience less pressure regarding attendance at work. Lastly, we find that teachers with full-time contracts (*employment status)* have significantly lower absence duration and frequency. These findings support the conclusions of Brown (1999) and Zeytinoglu et al. (2004), who highlighted that individuals with full-time contracts exhibit lower absenteeism levels compared to part-time contracts. This that full-time employment fosters stronger organisational attachment and accountability (Luke et al., 2019), thereby reducing absenteeism.

6.1.3 Conceptual Alignment: Literature and Study 1

Some antecedents were supported by the literature and identified in Study 1 but could not be tested in Study 2 due to data limitations. Despite the lack of quantitative validation, these factors reflect both theoretical grounding and experience, thus highlighting the need for future quantitative research.

In line with the findings from literature (Ehrenberg et al., 1991; Gerstenfeld, 1969; Wonders, 2021; Zuba & Schneider, 2013), Study 1 found that having *children* greatly impact teacher absenteeism. Highlighting that teachers with childcare responsibilities exhibit higher absenteeism rates, often due to needing flexibility or illness of children. Similarly, *health-status* was highlighted as a key predictors, aligning with numerous prior studies that found a direct relationship with absenteeism (Dwomoh & Moses, 2020; Goldberg & Waldman, 2000; Harrison & Martocchio, 1998; Melander et al., 2022). Study 1 also finds prominent job-related antecedents. *Job satisfaction* and *working environment/conditions* were found to influence teacher absenteeism, due to limited resources and excessive administrative tasks. These findings are consistent with existing literature (Green, 2014; Harrison & Martocchio, 1998; Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991; Shapira-Lishchinsky & Rosenblatt, 2010a). Study 1 finds that *autonomy* has consistently been decreased in Dutch secondary education due to the government forcing schools to spend a certain amount of time on specific tasks. Thereby, impacting teacher absenteeism, consistent with prior research (Bridges & Hallinan, 1978; Collie, 2021; Peng & Guo, 2022). Additionally, *stress* was frequently cited in Study 1, particularly when associated with workload pressures and behavioural challenges. This aligns with numerous, who found a direct link between stress and teacher absenteeism (Gaziel, 2004; Green, 2014; Miller et al., 2008). In line with this, negative s*tudent behaviour* was found to be a critical factor of creating stress, ultimately leading to teacher absenteeism. This aligns with previous research by Ismail (2023). Implying that negative student behaviour is not only a classroom concern but also a significant stressor that contributes to teacher absenteeism. Highlighting the need for schools to provide teachers with strategies to manage student conduct.

In addition, a positive, supportive, or servant *leadership style* was found to influence a teacher's commitment to attendance, aligned with research that linked these leadership styles with reduced absenteeism (Bowen, 2009; Bradley et al., 2007; Jung & Takeuchi, 2010). Underscoring that this leadership style fosters a stronger sense of commitment, support, and engagement, leading to fewer absences. Lastly, Study 1 also found that *additional responsibilities* (e.g. mentorship tasks) beyond regular teaching activities contributes to teacher absenteeism. Particularly among younger, less experienced teacher who may already struggle with core teaching tasks. These findings are in line with previous research from Kremer et al. (2005) and Usman et al. (2007). This implies that supplementary roles should be carefully assigned to teachers, addressing the importance of considering experience of teachers. These findings are consistent across literature, addressing their conceptual robustness, although they were not quantitatively tested in Study 2. Future research could integrate these antecedents into statistical models.

6.1.4 Empirical Convergence: Study 1 and Study 2

In addition to confirming established antecedents, both Study 1 and Study 2 revealed new teacher absenteeism antecedents: *employment at multiple schools* and *school type*. These are, to the best of our knowledge, not addressed in the current literature but emerged consistently across both studies. Consequently, we find support Hypotheses 10 and 15.

First of all, *employment at multiple schools* emerged as an potential influential antecedent in Study 1. This is supported by Study 2, which found that only male teachers working across multiple schools exhibited significantly higher absence duration compared to male teacher employed at one school. For absence frequency no relationship was found. A potential reasons could be that working a multiple locations increases complexity and logistical burden, impacting the absenteeism length rather than how often absences occur. Differences in *school type* also emerged in both studies. Study 1 briefly suggested that school type might influence absenteeims. Study 2 found a significant difference in absence duration between track-specific and vmbo schools, and a marginal difference in frequency between multi-level and vmbo schools. In Study 1 it was highlighted that practical education experiences higher absenteeism level due to large influx of new teachers and greater financial stability. Whereas this differences was not confirmed by the quantitative results. To fully validate these differences, the research should be extended to the whole geographical area of the Netherlands. Furthermore, this research finds no evidence that teacher function (LB, LC, and LD) influences absenteeism. This antecedent was not identified in Study 1, and Study 2 similarly found no significant relationship. Lastly, this research found no evidence between the relationship of teachers fulfilling *temporary or replacement* vacancies and absenteeism. This variable was not addressed in the literature or found in Study 1. Also, Study 2 found no evidence.

6.1.5 Conflicting Evidence and Discrepancies between Studies and Literature

Several antecedents showed inconsistencies across the literature, Study 1, and Study 2– either in significance or in its direction. The discrepancies between Study 1 and Study 2, are potentially due to the fact that many interviewees suggested certain patterns based on their experiences. Although they had not conducted a systematic analysis of their HR data comparable to that of Study 2, implying that some organisational assumptions about absenteeism may lack empirical validation and could benefit from data-driven insights. As Smulders (1984) noted, such discrepancies are common in behavioural research and often arise from methodological, contextual, or sample variations.

First of all, *age* presented a clear contradiction. The findings for absence duration are contrary to the results from Study 1 and literature, who found a negative relationship (Harrison & Martocchio, 1998; Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991).

Implying that older teachers have significantly less absence duration compared to younger teachers. Whereas no significant relationship was found between *age* and absence frequency. The reason for this discrepancies is that in Study 1 it was found that younger teachers have higher absenteeism due to having children, needing flexibility and some are following a study alongside teaching practices. Nevertheless, in Study 1 it was also indicated that older teacher struggle with digitalisation and dislike increasing administrative tasks, which aligns with Study 2. No relationship was found between *menopause* and teacher absenteeism in Study 2, contrary to Study 1 and past research (Geukes et al., 2016; O'Neill et al., 2023; Verdonk et al., 2022). This antecedent was incorporated as a rough proxy in Study 2, suggesting that all female employees in the menopause age range suffered from menopausal symptoms. As a result, it may have oversimplified a complex phenomenon, highlighting the need for more precise measurement in future research.

The negative relationship between *tenure* and absenteeism contradicts earlier findings that longer tenure is associated to higher absenteeism (Harrison & Martocchio, 1998; Scott & Wimbush, 1991). Instead this study find that teachers with longer tenure experience lower absence duration and frequency, possibly due to stronger organisational attachment, which may contribute to lower absenteeism. While Ann-Kristina and Nielsen (2008) found a significant positive correlation between employees' absence frequency and the absenteeism frequency of their managers. This study found no evidence to suggest that *managerial absence* predicts teacher absenteeism, in line with the findings from Duff et al. (2015). Moreover, Study 1 did not view it as relevant, implying that managerial absence is not associated with teacher absenteeism. Additionally, contrary to previous research (Knoster, 2016; Pfeifer, 2010; Scott & Wimbush, 1991), our findings show no relationships between *salary* and teacher absenteeism. Implying that salary does not affect teacher absenteeism. An explanation is that Dutch teachers have above average salaries (see Study 1) and follow fixed salary scales, making their minimum and maximum salary levels predetermined and widely known.[21]

[21] Rijksoverheid. (2024). *Scales, steps and rewards in secondary education*. https://www.rijksoverheid.nl/onderwerpen/werken-in-het-onderwijs/vraag-en-antwoord/wat-verdien-ik-als-leraar-in-het-voortgezet-onderwijs

Furthermore, Study 1 finds discrepancies regarding the directionality of relationships. While literature suggests that experienced teachers are less absent (Asiyai, 2017; Ost & Schiman, 2017; Speas, 2010), Study 1 finds that starting teachers are more frequently absent as they lack established resources and support. A similar discrepancy arises regarding the influence of monitoring and inspection. Our findings suggest that increased monitoring and inspection contributes to increased stress levels and reduced well-being, ultimately contributing to higher absenteeism. Thus, suggesting a positive relationship. On the other hand, the literature reveals that increased monitoring and inspection reduce absenteeism, hence suggesting a negative relationship (Cilliers et al., 2016; Muralidharan et al., 2017; Rogers et al., 2004). This discrepancy is likely since the studies are conducted in African and Asian school contexts, where enhance monitoring is often necessary. These antecedents are not statistically tested in Study 2, since no data was available. Future research should incorporate these variables to enable quantitative analysis and clarify their effects on absenteeism.

Lastly, Study 1 addressed briefly that *Municipality* (town or village) in which the school is located may impact teacher absenteeism. However, this is not supported by Study 2. This suggests that while participants in Study 1 perceived municipality as a potential factor influencing absenteeism based on their personal experiences or observations, Study 2 found no statistical evidence to support this claim for either absence duration or frequency. The finding from Study 1 lacks empirical grounding and highlights the importance of complementing these insights with data-driven analysis. Implying that municipality does not affect teacher absenteeism.

6.1.6 Uniquely Identified Antecedents in One Study

To the best of our knowledge, this research is the first to show certain antecedents that have not been highlighted in prior research. Study 1 found antecedents that lack supportive literature and could not be tested tested in Study 2 due to data availability. These include individual and contextual variables such as *perfectionism*, *busy social life*, *digitalisation*, *flexibility*, *following a study*, *unfavourable schedules*, *hierarchical structures*, *subject taught*, *tight labour and housing markets*, *parental influence*, *social safety*, and *negative public opinion*. While not statistically validated, their presence in interviews highlights the nuanced and situational nature of absenteeism and the value of qualitative approaches in surfacing new antecedents. To strengthen the empirical foundation of these findings, future research should aim to incorporate these variables into quantitative analyses to assess their significance.

Furthermore, Study 2 identified some significant antecedents not discussed in Study 1 or literature, but were included in the analysis due to data availability. As a result, our findings support Hypotheses 8 and 11. These include *FTE in team* (negatively associated with both duration and frequency) and *mangerial age* (negatively associated with frequency). This implies that larger teams may offer more peer support or workload flexibility, while older managers may foster more stable working environments. Nonetheless, these organisational variables deserve further attention in future research.

6.2 THEORETICAL CONTRIBUTIONS

This study makes several important contribution to the theoretical understanding of teacher absenteeism, particularly within the context of Dutch secondary education. Drawing on a multi-method sequential research design, this research contributes to the literature by confirming established antecedents, highlighting inconsistencies, and identifying new and context-specific antecedents.

First, our findings add to an improved understanding of teacher absenteeism by validating key existing antecedents in a new national context. While previous studies have extensively examined factors influencing absenteeism and its impact (Harrison & Martocchio, 1998; Knoster, 2016; Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991), research in the Dutch secondary education was lacking. By confirming several known antecedents– such as gender, prior absenteeism, working hours, and team-level absenteeism– within the Dutch secondary educational setting, this research demonstrates cross-contextual robustness.

Second, this research expands absenteeism literature by identifying school-specific and organisational antecedents that have not been highlighted by previous literature. Antecedents– such as FTE in team, school type, and managerial age emerged as significant predictors in the Dutch secondary educational contexts. These findings reflect a more comprehensive understanding of absenteeism that extends traditional absenteeism research (Biron & Bamberger, 2010; Harrison & Martocchio, 1998; Knoster, 2016; Rosenblatt & Shirom, 2005; Scott & Wimbush, 1991). Highlighting that organisational design and institutional context are important when researching absenteeism. The extensive set of antecedents used in this research, aligns with a growing trend to incorporate big data and analytics in HR research (Biron & Bamberger, 2010; Knoster, 2016).

Third, this research shows inconsistencies in how certain antecedents operate differently across different empirical methods and contexts. Our findings highlight discrepancies in both the significance of relationship (e.g., managerial absence and salary) and their directionality (e.g., age and tenure). These findings show the importance of examining absenteeism within specific context, as they reveal limited generalisability across settings. This aligns with Smulders (1984), who emphasised that absenteeism research is often difficult to generalise due to incomplete data and discrepancies in the operationalisation of absenteeism. Implying that context-specific research is necessary to get a more comprehensive understanding absence behaviour and should be applied with caution across different contexts.

Furthermore, this research reveals that some antecedents were significantly related with absence duration but not with absence frequency (e.g., age). Whereas other antecedents significantly influenced absence frequency but not duration (e.g., managerial age). This indicates that antecedents driving how long teachers are absent may differ from those influencing how often they are absent. Thereby, improving our understanding of teacher absenteeism. A potential reason for this discrepancy is that teachers may take frequent short absences to cope with manageable, day-to-day stressors, whereas longer absence are typically triggered by more serious challenges. These discrepancies highlight the multi-faceted nature of absenteeism and show the importance of examining it from both perspectives. Focusing on only one construct could lead to an incomplete understanding of absence behaviour. Nonetheless, our findings shows that the directionality of the relationship remained consistent when significance across both outcomes.

Finally, the integration of ML in Study 3 offers a novel methodological understanding of teacher absenteeism. While predictive analytics have been increasingly applied in HR research (Biron & Bamberger, 2010), few studies have used such techniques to classify types of absenteeism in educational contexts. Moreover, to the best of my knowledge, no prior study has employed a two-step approach in which statistically validated predictors from regression analysis (Study 2) are first identified and then used as input features for ML models (Study 3). This structured method strengthens theoretical justification for the model’s variables but also increases transparency ('black box'), which is one of the most critiques of utilising ML models (Rudin & Radin, 2019). By empirically grounding antecedents, this research enhances the interpretability. The Naïve Bayes model developed in Study 3 achieved a classification accuracy of 90.7%, outperforming prior models such as Ajmi (2019), which reached 83.3%, and aligning with research by Gayathri (2018). These results underscore the value of combining theory-driven modelling with advanced analytics to operationalise complex behavioural constructs like absenteeism.

6.3 PRACTICAL CONTRIBUTIONS

From a practical perspective, the research offers valuable insights for educational institutions and HR professionals to understand, predict, and manage teacher absenteeism more effectively. Our findings provide both theoretically grounded and data-driven contributions that can inform real-world absenteeism interventions in the Dutch secondary education context.

First, the consistent identification of significant predictors– such as gender, working hours, team-level absenteeism, and prior absenteeism– enables schools to better understand which teachers may be at greater risk of absenteeism. While some antecedents (e.g., age and gender) cannot be changed and are difficult to tackle as organisations, they can still inform the development of preventive absence strategies. For example, teachers with a history of high absenteeism could be prioritised for early support, such as coaching, mentoring, or flexible workload adjustments. Similarly, schools with smaller team sizes may offer increased peer support and monitor attandance norms more closely to mitiage a potential 'absence culture'. Similarly, our findings address that teachers under older managers had lower absenteeism frequency, implying that older managers may offer a more stable experienced leadership. Therefore, schools might consider team structuring as part of broader HR stategies. These practical insights potentially reduce absenteeism, eventually leading to sustainable employability. Aligning with broader HR management goals in education (Green, 2014; Jaarsveld & Keyser, 2018). Moreover, our findings show that school specific antecedents (employment at multiple schools and school type) influence absenteeism. Consequently, school should carefully consider teachers movement between institutions to reduce strain.

Furthermore, our research addresses the practical need from educational institutions to move from current traditional descriptive analytics to predictive absenteeism analytics (Hoppenbrouwer & Bouma, 2021). Our findings show how certain antecedents influence teacher absenteeism, thereby providing a more comprehensive understanding of teacher absenteeism. Moreover, Study 3 demonstrates the practical utilisation of ML models in educational HR analytics. Our findings demonstrate, with an accuracy of 90.7%, that individual absence duration categories can be predicted. By predicting short-, middle, and long-term absenteeism based on historical data, educational institutions can optimise proactive absence management strategies, improve resource allocation, workload distribution, and forecast absenteeism trends. This predictive capability offers substantial value for schools, especially due to growing teacher shortages [22]. More specifically, provide high-risk profiles with support (e.g., coaching, monitoring, and workload adjustments). Finally, these insights hold commercial potential for Infotopics by offering more comprehensive predictive HR analytics solutions to educational institutions and a broader range of clients.

6.4 LIMITATIONS AND FUTURE RESEARCH

This research has certain limitations that should be acknowledged. Firstly, while a broad range of personal, job-related, and socio-economic antecedents have been analysed in this research. There remain several antecedents identified in the literature and Study 1 that were not covered in Study 2 due to data availability constraints. Antecedents such as job satisfaction, stress, autonomy, experience, and children have been recognised as important factors influencing absenteeism, but were not included in the quantitative analysis (Bradley et al., 2007; Cohen & Golan, 2007; Harrison & Martocchio, 1998; Porter & Steers, 1973; Rosenblatt & Shirom, 2005; Shapira-Lishchinsky & Rosenblatt, 2010b). Similarly, Study 1 found that unfavourable schedules, teaching subject, and the number of location changes may impact absenteeism, yet these factors were also not quantitatively analysed. Future research should seek to analyse these antecedents by integrating a more comprehensive dataset, enabling complete and statistical validation of the findings from Study 1. Second, is the generalisability of the findings (Smulders, 1984). This study is specific to Dutch secondary education, and while some findings may apply to other contexts, patterns may differ across primary education, higher education and international settings. Future research should

[22] Dutch Ministry of Education, Culture, and Science (2023). *Trendrapportage Arbeidsmarkt Leraren po, vo en mbo 2023*. https://www.aob.nl/assets/Bijlage-3.1-Trendrapportage-Arbeidsmarkt-Leraren-2023.pdf

examine absenteeism in different educational sector and across multiple countries to improve generalisability.

The third limitation is external validity. Although the current dataset provided valuable insights, a larger dataset covering multiple educational institutions would improve the external validity of both Study 2 and Study 3. Particularly for Study 3, the class imbalance in the dataset resulted in lower classification accuracy for long-term absences due to limited number of instances. Future research could optimise model performance by addressing class imbalance, incorporate additional predictors and real-time data, thereby improving the accuracy of the Naïve Bayes model.

Moreover, while Study 3 demonstrates the practical potential of ML in predicting absence duration categories, it also raises critical questions regarding its implementation within educational institutions. Consequently, future research must address the ethical implication and accountability associated with predictive absence analytics in schools. For instance, what happens when a teacher is misclassified as high-risk for long-term absenteeism. Therefore, future studies should not only improve predictive accuracy but also investigate how such models can be implemented responsibly, ensuring transparency and ethical accountability.

Finally, future research should go beyond identifying antecedents of absenteeism by conducting experimental research within schools. This enables to capture how teachers experience and respond to certain absence management strategies. Moreover, it provides evidence on what works to reduce absenteeism, and thus offer stronger guidance for school principals.

6.5 CONCLUSION

This first part of this research aimed to identify key antecedents influencing teacher absenteeism in Dutch secondary education. The research question was: *"What are the domain-specific antecedents of teacher absenteeism in Dutch secondary education?"* Our findings identifies an extensive amount of antecedents in Dutch secondary education, highlighting both established and novel antecedents. While Study 1 confirmed established antecedents (e.g. children, autonomy and stress) and novel antecedents (e.g., less favourable schedule and parental influence), these antecedents require further statistical validation to assess their significance. Interviewees suggested certain absenteeism patterns based on their experiences, while this was not quantitatively grounded and lacking empirical validation, as most acknowledged during the interviews. In addition, Study 2 revealed that certain antecedent only emerge in relation to either absence duration or frequency (e.g., age and contract type). Consequently, the 'key' antecedents that consistently influence teacher absenteeism across duration and frequency seem to be the most important. These include: *gender, prior absenteeism, tenure, working hours, team-level absenteeism,* and *employment status*. Additionally, *school type* and *FTE in team* were identified as novel antecedents across both perspectives.

The second part aimed to develop a predictive model for classifying absence duration categories. Consequently the research question was: *"How accurately can machine learning classify teacher absence duration categories?"* Our predictive ML model (Naïve Bayes) demonstrated an accuracy of 90.7% in classifying individual teacher absence duration categories. These findings contribute to a deeper understanding of teacher absenteeism in Dutch secondary education and offers valuable insights for proactive absence management strategies.

REFERENCES

Ajmi, S. Q. (2019). Predicting Absenteeism at Work Using Machine Learning Algorithms. *Muthanna Journal of Pure Science*, *7*(1). https://doi.org/10.52113/2/07.01.2020/1-12

Allebeck, P., & Mastekaasa, A. (2004). Chapter 5. Risk factors for sick leave - general studies. *Scandinavian Journal of Public Health*, *32*(63), 48-108. https://doi.org/https://doi.org/10.1080/14034950410021853

Alwosheel, A., Cranenburgh, S. v., & Chorus, C. G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, *28*, 167-182. https://doi.org/https://doi.org/10.1016/j.jocm.2018.07.002

Ann-Kristina, & Nielsen, L. (2008). Determinants of Absenteeism in a Public Sector Perspective. *The International Journal of Human Resource Management*. https://pure.au.dk/ws/portalfiles/portal/7453/Nielsen_2009_e

Arai, M., & Thoursie, P. S. (2005). Incentives and Selection in Cyclical Absenteeism. *Labour Economics*, *12*(2), 269-280. https://doi.org/https://doi.org/10.1016/j.labeco.2003.11.009

Asamoah, M. K. (2014). Re-examination of the limitations associated with correlational research. *Journal of Educational Research and Reviews*, *2*(4), 45-52. http://www.sciencewebpublishing.net/jerr/archive/2014/July/pdf/Asamoah.pdf

Asiyai, R. I. (2017). Exploring Teacher Absenteeism Dilemma in Public Secondary Schools in Delta State, Nigeria. *Issues and Ideas in Education*, *5*(1), 73-86. https://doi.org/https://doi.org/10.15415/iie.2017.51005

Astivia, O. L. O., & Zumbo, B. D. (2019). Heteroskedasticity in multiple regression analysis: What it is, how to detect it and how to solve it with applications in R and SPSS. *Journal of Modern Applied Statistical Methods*, *24*(1).

Baker, S. E., & Edwards, R. (2012). How many qualitative interviews is enough? *National Centre for Research Methods Review Paper*. https://eprints.ncrm.ac.uk/id/eprint/2273/4/how_many_interviews.pdf

Banerjee, R., King, E. M., Orazem, P. F., & Paterno, E. M. (2012). Student and teacher attendence: The role of shared goods in reducing absenteeism. *Economics of Education Review*, *31*(5), 563-574. https://doi.org/https://doi.org/10.1016/j.econedurev.2012.04.002

Barmby, T., Nolan, M., & Winkelmann, R. (2001). Contracted Workdays and Absence. *The Manchester School*, *69*(3). https://doi.org/https://doi.org/10.1111/1467-9957.00247

Baydoun, M., Dumit, N. Y., & Daouk-Öyry, L. (2015). What do nurse managers say about nurses' sickness absenteeism? A new perspective. *Journal of Nursing Management*, 97-104. https://doi.org/https://doi.org/10.1111/jonm.12277

Belita, A., Mbindyo, P., & English, M. (2013). Absenteeism amongst health workers – developing a typology to support empiric work in low-income countries and characterizing reported associations. *Human Resources for Health*, *11*(1). https://doi.org/https://doi.org/10.1186/1478-4491-11-34

Bhushan, A. (2018). Teachers Autonomy: A Tool for Creating Learners. *University of Mumbai*. https://www.researchgate.net/publication/330411772_Teachers_Autonomy_A_Tool_for_Creating_Learners_Autonomy

Biron, M., & Bamberger, P. A. (2010). The impact of structural empowerment on individual well-being and performance: Taking agent preferences, self-efficacy and operational constraints into account. *Human Relations*, *63*(2), 163-191. https://doi.org/10.1177/0018726709337039

Borman, G. D., & Dowling, N. M. (2008). Teacher Attrition and Retention: A Meta-Analytic and Narrative Review of the Research. *Review of Educational Research*, *78*(3), 367-409. https://doi.org/https://doi.org/10.3102/0034654308321455

Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40. https://doi.org/10.3316/QRJ0902027

Bowers, A. J., & White, B. R. (2014). Do Principal Preparation and Teacher Qualifications Influence Different Types of School Growth Trajectories in Illinois?A Growth Mixture Model Analysis. *Journal of Educational Administration*, *52*(2), 705-736. https://doi.org/http://doi.org/10.1108/JEA-12-2012-0134

Box, G. E. P., & Cox, D. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243. https://doi.org/https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Bradley, S., Green, C., & Leeves, G. (2007). Worker absence and shrinking: Evidence from matched teacher-school data. *Labour Economics*, *14*(3), 319-334. https://doi.org/https://doi.org/10.1016/j.labeco.2006.05.002

Bridges, E. M., & Hallinan, M. T. (1978). Subunit Size, Work System Interdependence, and Employee Absenteeism. *Educational Administration Quarterly*, *14*(2), 24-42. https://doi.org/https://doi.org/10.1177/0013131X7801400203

Brinkmann, S., & Kvale, S. (2009). *InterViews: Learning the Craft of Qualitative Research Interviewing*. Sage Publications. https://psycnet.apa.org/record/2008-15512-000

Brooke, P. P., & Price, J. L. (1989). The determinants of employee absenteeism: An empirical test of a causal model. *Journal of Occupational Psychology*, *62*(1), 1-19. https://doi.org/https://doi.org/10.1111/j.2044-8325.1989.tb00473.x

Brown, S. (1999). Worker absenteeism and overtime bans. *Applied Economics*, *31*(2), 165. https://doi.org/10.1080/000368499324390

Carrizosa, L. M. G., & Witte, K. D. (2024). Teacher absenteeism: a conceptual model developed from a systematic literature review. *Cambridge Journal of Education*, *54*(4), 561-588. https://doi.org/https://doi.org/10.1080/0305764X.2024.2397366

Cheloha, R. S., & Farr, J. L. (1980). Absenteeism, job involvement, and job satisfaction in an organizational setting. *Journal of Applied Psychology*, *65*(4), 467-473. https://doi.org/https://doi.org/10.1037/0021-9010.65.4.467

Cilliers, J., Ibrahim, K., Leaver, C., Serneels, P., & Zeitlin, A. (2016). Pay for Locally Monitored Performance? A Welfare Analysis for Teacher Attendance in Ugandan Primary Schools. *SSRN Electronic Journal 167*, 69-90. https://doi.org/https://doi.org/10.2139/ssrn.2822657

Cohen, A., & Golan, R. (2007). Predicting absenteeism and turnover intentions by past absenteeism and work attitudes: An empirical examination of female employees in long term nursing care facilities. *Career Development International*, *12*(5), 416-432. https://doi.org/https://doi.org/10.1108/13620430710773745

Collie, R. J. (2021). COVID-19 and Teachers' Somatic Burden, Stress, and Emotional Exhaustion: Examining the Role of Principal Leadership and Workplace Buoyancy. *AREA Open 7*. https://doi.org/10.1177/2332858420986187

Consul, P. C., & Famoye, F. (2007). Generalized poisson regression model. *Communications in Statistics - Theory and Methods*, *21*(1), 89-109. https://doi.org/https://doi.org/10.1080/03610929208830766

Coxe, S., West, S. G., & Aiken, L. S. (2008). The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. *Journal of Personality*

Assessment, *91*(2), 121-136. https://doi.org/https://doi.org/10.1080/00223890802634175

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* (3rd ed.). SAGE Publications. https://www.ucg.ac.me/skladiste/blog_609332/objava_105202/fajlovi/Creswell.pdf

Cunningham, G. O. (2019). Teacher absenteeism and retention: the predictive relationship between principal leadership style and perceived ethic of care. *International Journal of Humanities and Social Science*, *9*(2). https://doi.org/https://doi.org/10.30845/ijhss.v9n2p1

Deakin, H., & Wakefield, K. (2013). Skype interviewing: reflections of two PhD researchers. *Qualitative Research 14*(5), 603-616. https://doi.org/https://doi.org/10.1177/1468794113488126

Dismuke, C., & Lindrooth, R. (2006). *Ordinary least squares. Methods and designs for outcomes research* (Vol. 93).

Drago, R., & Wooden, M. (1992). The determinants of Labor Absence: Economic Factors and Workgroup Norms across Countries. *ILR Review*, *45*(4), 764-778. https://doi.org/https://doi.org/10.1177/001979399204500411

Duff, A. J., Podolsky, M., Biron, M., & Chan, C. C. A. (2015). The interactive effect of team and manager absence on employee absence: A multilevel field of study. *Journal of Occupational & Organizational Psychology*, *88*(1), 61-79. https://doi.org/https://doi.org/10.1111/joop.12078

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*, *6*(3), 241-252. https://doi.org/https://doi.org/10.1080/00401706.1964.10490181

Dwomoh, E., & Moses, M. O. (2020). Job-related activity patterns, health status and absenteeism-related factors of star-rated hotels staff. *European Journal of Tourism Hospitality and Recreation*, *10*(3), 274-285. https://doi.org/https://doi.org/10.2478/ejthr-2020-0024

Edge, M. D. (2019). *Statistical Thinking from Scratch*. Oxford University Press.

Edwards, V. L. (2014). Examining absenteeism in the public and non-profit sectors. *International Journal of Organization Theory and Behavior*, *17*(3), 293-310. https://doi.org/10.1108/IJOTB-17-03-2014-B002

Ehrenberg, R. G., Ehrenberg, R. A., Rees, D. I., & Ehrenberg, E. L. (1991). School district leave policies, teacher absenteeism, and student achievement. *Journal of Human Resources*, *26*, 72-105. http://www.jstor.org/stable/145717

Farrell, D., & Stamm, C. L. (1988). Meta-analysis of the correlates of employee absence. *Human Relations*, *41*(3), 211-227. https://doi.org/https://doi.org/10.1177/001872678804100302

Fernandes, F. T., & Filho, A. D. P. C. (2021). Prediction of absenteeism in public schools teachers with machine learning. *Revista de saude publica*. https://doi.org/https://doi.org/10.11606/s1518-8787.2021055002677

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). SAGE Publications. https://www.sagepub.com/sites/default/files/upm-binaries/52063_00_Field_4e_SPSS_Prelims.pdf

Garcia, R. L. (1987). Sick-Time Usage by Management and Professional Employees in the Public Sector. *Review of Public Personnel Administration*, *7*(3), 45-59. https://doi.org/https://doi.org/10.1177/0734371X8700700305

Garrison, K. R., & Muchinsky, P. M. (1977). Attitudinal and biographical predictors of incidental absenteeism. *Journal of Vocational Behavior*, *10*(2), 221-230. https://doi.org/https://doi.org/10.1016/0001-8791(77)90059-8

Gayathri, T. (2018). Data mining of Absentee data to increase productivity. *International Journal of Engineering and Techniques*, *4*(3). http://www.ijetjournal.org/Volume4/Issue3/IJET-V4I3P78.pdf

Gaziel, H. M. (2004). Predictors of absenteeism among primary school teachers. *Social Psychology of Education*, *7*, 421-434. https://doi.org/10.1007/s11218-004-5232-z

Gellatly, I. R. (1995). Individual and Group Determinants of Employee Absenteeism: Test of a Causal Model. *Journal of Organizational Behavior*, *16*(5), 469-485. https://doi.org/http://www.jstor.org/stable/2488600

Gerstenfeld, A. (1969). Employee absenteeism: New insights: Data reveal external factors. *Business Horizons*, *12*(5), 51-57. https://doi.org/https://doi.org/10.1016/0007-6813(69)90041-X

Geukes, M., Aalst, M. P. v., Robroek, S. J. W., Laven, J. S. E., & Oosterhof, H. (2016). The impact of menopause on work ability in women with severe menopausal symptoms. *Maturitas*, *90*, 3-8. https://doi.org/https://doi.org/10.1016/j.maturitas.2016.05.001

Geurts, S. A. E., Buunk, B. P., & Schaufeli, W. B. (1994). Social comparisons and absenteeism: a structural modeling approach. *Journal of Applied Social Psychology*, *24*(21), 1871-1890. https://doi.org/https://doi.org/10.1111/j.1559-1816.1994.tb00565.x

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, *16*(1), 15-31. https://doi.org/https://doi.org/10.1177/1094428112452151

Goldberg, C. B., & Waldman, D. (2000). Modeling employee absenteeism: Testing alternative measures and mediated effects based on job satisfaction. *Journal of Organizational Behavior*, *21*(6), 665-676. https://doi.org/10.1002/1099-1379(200009)21:63.0.CO;2-P

Gosselin, E., Lemyre, L., & Corneil, W. (2013). Presenteeism and absenteeism: Differentiated understanding of related phenomena. *Journal of Occupational Health Psychology*, *18*(1), 75-86. https://doi.org/https://doi.org/10.1037/a0030932

Green, G. R. (2014). Study to investigate self-reported teacher absenteeism and desire to leave teaching as they relate to teacher-reported teaching satisfaction, job-related stress, symptoms of depression, irrational beliefs and self-efficacy. *CUNY Academic Works*. https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1318&context=gc_etds

Hackett, R. D. (1990). Age, tenure, and employee absenteeism. *Human Relations*, *43*(7), 601-619. https://doi.org/https://doi.org/10.1177/001872679004300701

Hangcheng, l. (2015). Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance. *Virginia Commonwealth University*.

Harrison, D. A., & Martocchio, J. J. (1998). Time for Absenteeism: A 20-Year Review of Origins, Offshoots, and Outcome. *Journal of Management*, *24*(3), 305-350. http://ereserve.library.utah.edu/Annual/MGT/7800/Brief/harrison-1.pdf

Hassan, S., Wright, B. E., & Yuki, G. (2014). Does Ethical Leadership Matter in Government? Effects on Organizational Commitment, Absenteeism, and Willingness to Report Ethical Problems. *Public Administration Review*, *74*(3). https://doi.org/10.1111/puar.12216

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, *39*(4), 709-722. https://doi.org/https://doi.org/10.3758/BF03192961

Heinzl, H., & Mittlböck, M. (2003). Pseudo R-squared measures for Poisson regression models with over- or underdispersion. *Computational Statistics & Data Analysis*, *44*(2), 253-271. https://doi.org/https://doi.org/10.1016/S0167-9473(03)00062-8

Hendriks, M. (2024). Werkgeluk en sociale veiligheid als wapens tegen verzuim. *Erasmus Happiness Economics Research Organisation*, 56-68.

Henseler, J., Schuberth, F., & Yu, X. (2024). Regression Analysis - technique lecture. https://canvas.utwente.nl/courses/14129/assignments/syllabus

Hoppenbrouwer, X., & Bouma, J. (2021). *Thema 'verzuim'*.

Imants, J., & Zoelen, A. v. (1995). Teachers' Sickness Absence in Primary Schools, School Climate and Teachers' Sense of Efficacy. *School Organisation*, *15*, 77-86. https://doi.org/10.1080/02601369550038372

Ismail, Z. (2023). Identify the Level of Emotional Stress Factors for Workload, Lack of Resources, Administrative Management, Time Constraints and Negative Student Behavior and Disciplinary Problems Among Teachers at Sekolah Menengah Kebangsaan Pangkal Meleret, Machang, Kelantan. *Asian Journal of Vocational Education and Humanities*, *4*(2), 31-39. https://doi.org/https://doi.org/10.53797/ajvah.v4i2.4.2023

Iverson, R. D., Olekalns, M., & Erwin, P. J. (1998). Affectivity, Organizational Stressors, and Absenteeism: A Causal Model of Burnout and Its Consequences. *Journal of Vocational Behavior*, *52*(1), 1-23. https://doi.org/https://doi.org/10.1006/jvbe.1996.1556

Jaarsveld, D. v., & Keyser, E. (2018). The Moderating Role of Job Satisfaction on Workplace Absenteeism and Substance use amongst the Employees at a Power Utility in Mpumalanga. *Journal of Economics and Behavioral Studies*, *10*(2), 219-230.

Jacob, B. A. (2010). The Effect of Employment Protection on Worker Effort: Evidence from Public Schooling. *National Bureau of Economic Research*. https://doi.org/https://doi.org/10.3386/w15655

Jean-Francois, T. (2023). Teachers' perspective on accountability: A comparative case study in primary rural schools of Afghanistan and Pakistan. *Research in Comparative and International Education*, *18*(4), 569-599. https://doi.org/https://doi.org/10.1177/17454999231186878

Johns, G. (2008). Absenteeism and presenteeism: Not at work or not working well. *SAGE Publications*. https://doi.org/10.4135/9781849200448.n10

Johns, G. (2010). Presenteeism in the workplace: A review and research agenda. *Journal of Organizational Behavior*, *31*(4), 519-542. https://doi.org/ https://doi.org/10.1002/job.630

Johnston, M. P. (2014). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries*, *3*(3), 619-626. https://www.researchgate.net/publication/294718657_Secondary_Data_Analysis_A_Method_of_Which_the_Time_has_Come

Judge, T. A., Martocchio, J., & Thoresen, C. J. (1997). Five-Factor Model of Personality and Employee Absence. *Journal of Applied Psychology*, *82*(5), 745-755. https://doi.org/https://doi.org/10.1037/0021-9010.82.5.745

Jung, Y., & Takeuchi, N. (2010). Performance implications for the relationships among top management leadership, organizational culture, and appraisal practice: Testing two theory-based models of organizational learning theory in Japan. *The International Journal of Human Resource Management*, *21*(11), 1931-1950. https://doi.org/10.1080/09585192.2010.505093

Kamphuis, R. (2018). Using Stepwise Regression Techniques to Shortlist the Number of Antecedents of Employee
Absenteeism. *University of Twente*.

Kanungo, R. N. (1982). Measurement of job and work involvement. *Journal of Applied Psychology*, *67*(3), 341-349. https://doi.org/https://doi.org/10.1037/0021-9010.67.3.341

Keller, R. T. (1983). Predicting absenteeism from prior absenteeism, attitudinal factors, and nonattitudinal factors. *Journal of Applied Psychology*, 536-540. https://doi.org/https://doi.org/10.1037/0021-9010.68.3.536

Knoster, K. C. (2016). Strategies for Addressing Student and Teacher Absenteeism: A Literature Review. *North Central Comprehensive Center*. https://compcenternetwork.org/sites/default/files/archive/nccc_absenteeism_report-2.pdf

Kolk, H. v. d. (2021). *Intro R for BMS Bachelor programs*. https://rhelpdesk.utwente.io/1-introduction.html

Kolk, H. v. d. (2023). Microlecture – Testing for Normality: Shapiro-Wilk Test.

Kremer, M., Chaudhury, N., Rogers, F. H., & Muralidharan, K. (2005). Teacher Absence in India: A Snapshot. *Journal of the European Economic Association*, *3*(2), 658-667. https://doi.org/https://doi.org/10.1162/1542476054473143

Kump, B., & Dahlke, J. (2024). *Lecture on Qualitative Research and Business Skills*. University of Twente. https://canvas.utwente.nl/courses/14128/files/4327656?module_item_id=498923

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology*, *70*(2), 144-156. https://doi.org/https://doi.org/10.4097/kjae.2017.70.2.144

Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer Publishing Company.

Lenth, R. V. (2025). Estimated Marginal Means, aka Least-Squares Means. *Package 'emmeans'*. https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Leuven, E. (2008). Causes and consequences of teacher absenteeism. *Dutch Research Council*. https://www.nwo.nl/en/projects/451-04-003-0

Luke, M. A., Carnelley, K. B., & Sedikides, C. (2019). Attachments in the workplace: How attachment security in the workplace benefits the organisation. *European Journal of Social Psychology*, *50*(5), 1046-1064. https://doi.org/10.1002/ejsp.2652

Lyon, J. D., & Tsai, C.-L. (1996). A Comparison of Tests for Heteroscedasticity. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *45*(3), 337–349. https://doi.org/https://doi.org/10.2307/2988471

Magnani, G., & Gioia, D. (2023). Using the Gioia Methodology in international business and entrepreneurship research. *International Business Review*, *32*(2). https://doi.org/https://doi.org/10.1016/j.ibusrev.2022.102097

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research*, *9*(1), 381-386. https://doi.org/10.21275/ART20203995

Martocchio, J. J. (1989). Age-related differences in employee absenteeism: a meta-analysis. *Psychology and Aging*, *4*(4), 409-414. https://doi.org/https://doi.org/10.1037/0882-7974.4.4.409

Mbachu, H. I., Nduka, E. C., & Nja, M. E. (2012). Designing a Pseudo R-Squared Goodness-of-Fit Measure

in Generalized Linear Models. *Journal of Mathematics Research 4*(2), 148. https://doi.org/doi:10.5539/jmr.v4n2p148

Melander, K., Kortteisto, T., Hermanson, E., Kaltiala, R., Mäki-Kokkila, K., Kaila, M., & Kosola, S. (2022). The perceptions of different professionals on school absenteeism and the role of school health care: A focus group study conducted in Finland. *Plos One*, *17*(2). https://doi.org/https://doi.org/10.1371/journal.pone.0264259

Miller, R. T., Murnane, R. J., & Willett, J. B. (2008). Do Teacher Absences Impact Student Achievement? Longitudinal Evidence From One Urban School District. *Educational Evaluation and Policy Analysis*, *30*(2), 181-200. https://doi.org/https://doi.org/10.3102/0162373708318019

Montano, I. H., Marques, G., Alonso, S. G., López-Coronado, M., & Díez, I. d. l. T. (2020). Predicting Absenteeism and Temporary Disability Using Machine Learning: a Systematic Review and Analysis. *Systems-Level Quality Improvement*, *44*(162). https://doi.org/https://doi-org.ezproxy2.utwente.nl/10.1007/s10916-020-01626-2

Muralidharan, K., Das, J., Holla, A., & Mohpal, A. (2017). The fiscal cost of weak governance: Evidence from teacher absence in India. *Journal of Public Economics*, *145*, 116-135. https://doi.org/https://doi.org/10.1016/j.jpubeco.2016.11.005

Nedungadi, P., Mulki, K., & Raman3, R. (2017). Improving educational outcomes & reducing absenteeism at remote villages with mobile technology and WhatsAPP: Findings from rural India. *Education and Information Technologies*, *23§*(1), 113-127. https://doi.org/https://doi.org/10.1007/s10639-017-9588-z

Nicholson, S., Pauly, M. V., Polsky, D., Sharda, C., Szrek, H., & Berger, M. L. (2006). Measuring the effects of work loss on productivity with team productio. *Health Economics*, *15*, 111-123. https://doi.org/10.1002/hec.1052

Nielsen, K., & Daniels, K. (2016). The relationship between transformational leadership and follower sickness absence: the role of presenteeism. *Work & Amp Stress*, *30*(2), 193-208. https://doi.org/https://doi.org/10.1080/02678373.2016.1170736

O'Neill, M., Jones, V. F., & Reid, A. (2023). Impact of menopausal symptoms on work and careers: a cross-sectional study. *Occupational Medicine*, *73*(6), 332-338. https://doi.org/https://doi.org/10.1093/occmed/kqad078

Ommeren, J. N. v., & Gutiérrez-i-Puigarnau, E. (2011). Are workers with a long commute less productive? An empirical analysis of absenteeism. *Regional Science and Urban Economics*, *41*(1), 1-8. https://doi.org/https://doi.org/10.1016/j.regsciurbeco.2010.07.005

Ost, B., & Schiman, J. C. (2017). Workload and teacher absence. *Economics of Education Review*, *57*, 20-30. https://doi.org/https://doi.org/10.1016/j.econedurev.2017.01.002.

Östlund, U., Kidd, L., Wengström, Y., & Rowa-Dewar, N. (2011). Combining qualitative and quantitative research within mixed method research designs: a methodological review. *International Journal of Nursing Studies*, *48*(3), 369-383. https://doi.org/https://doi.org/10.1016/j.ijnurstu.2010.10.005

Owen, A. T. (2010). Leadership Practices That Influence Teacher Attendance in a Low and High Teacher Absentee School. *Georgia Southern University*. https://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=1326&context=etd

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 553-544. https://doi.org/https://doi.org/10.1007/s10488-013-0528-y

Peng, Y., & Guo, C. (2022). The Relationship between Teacher Autonomy and Mental Health in Primary and Secondary School Teachers: The Chain-Mediating Role of Teaching Efficacy and Job Satisfaction. *International Journal of Environmental Research and Public Health*, *19*(22). https://doi.org/https://doi.org/10.3390/ijerph192215021

Pfeifer, C. (2010). Impact of wages and job levels on worker absenteeism. *International Journal of Manpower*, *31*(1), 59-72. https://doi.org/https://doi.org/10.1108/01437721011031694

Podgursky, M. (2003). Fringe Benefits. *Education Next 3*. http://web.missouri.edu/~podgurskym/articles/files/fringe_benefits.pdf

Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin,*, *80*(2), 151-176. https://doi.org/https://doi.org/10.1037/h0034829

Possenriede, D., Hassink, W., & Plantenga, J. (2014). Does temporal and locational flexibility of work reduce absenteeism? *Utrecht School of Economics* https://www.uu.nl/sites/default/files/use_14-09.pdf

Price, H. E. (2012). Principal–Teacher Interactions: How Affective Relationships Shape Principal and Teacher Attitudes. *Educational Administration Quarterly*, *48*(1), 39-85. https://doi.org/https://doi.org/10.1177/0013161X11417126

Rentsch, J. R., & Steel, R. P. (2003). What does unit-level absence mean? Issues for future unit-level absence research. *Human Resource Management Review*, *13*(2), 185-202. https://doi.org/https://doi.org/10.1016/S1053-4822(03)00012-3

Robinson, C., & Schumacker, R. E. (2009). Interaction effects: centering, variance inflation factor, and interpretation issues. *Multiple linear regression viewpoints*, *35*(1), 6-11.

Rogers, F. H., López-Cálix, J., Chaudhury, N., Hammer, J., Córdoba, N., Kremer, M., & Muralidharan, K. (2004). Teacher absence and incentives in primary education. *World Bank and Inter–American Development Bank*, *2*, 136-162. https://assets.publishing.service.gov.uk/media/57a08cbbed915d622c001531/Rogers_etal_Teacherabsence_1_.pdf

Rogers, F. H., & Vegas, E. (2009). No more cutting class? Reducing teacher absence and providing incentives for performance. *World Bank*. https://www.researchgate.net/publication/48139828_No_more_cutting_class_Reducing_teacher_absence_and_providing_incentives_for_performance

Rosenblatt, Z., & Shirom, A. (2005). Predicting teacher absenteeism by personal background factors. *Journal of Educational Administration*, *43*, 209-225. https://doi.org/https://doi.org/10.1108/09578230510586597

Rudin, C., & Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition. *Harvard Data Science Review*, *1*(2). https://doi.org/https://doi.org/10.1162/99608f92.5a8a3a3d

Saenz-Armstrong, P. (2020). Roll Call 2020. *National Council on Teacher Quality*. https://www.nctq.org/pages/Roll-Call-2020

Schlagman, S., & Kvavilashvili, L. (2008). Involuntary autobiographical memories in and outside the laboratory: How different are they from voluntary autobiographical memories? *Journal of Memory & Cognition*, *36*(5), 920-932. https://doi.org/https://doi.org/10.3758/mc.36.5.920

Schultz, A. B., Chen, C.-Y., & Edington, D. W. (2009). The cost and impact of health conditions on presenteeism to employers: a review of the literature. *PharmacoEconomics*, *27*(5), 365-378. https://doi.org/https://doi.org/10.2165/00019053-200927050-00002

Scott, K. D., & Wimbush, J. C. (1991). Teacher Absenteeism in Secondary Education. *Educational Administration Quarterly*, *27*(4), 506-529. https://doi.org/https://doi.org/10.1177/0013161X91027004004

Senou, B. M. (2021). Contract Type and Teacher Absenteeism in Benin: The Role of Teacher's Supplemental Income. *African Economic Research Consortium*. https://publication.aercafricalibrary.org/handle/123456789/2216

Shapira-Lishchinsky, O., & Rosenblatt, Z. (2010a). School ethical climate and teachers' voluntary absence. *Journal of Educational Administration*, *48*(2), 164-181. https://doi.org/https://doi.org/10.1108/09578231011027833

Shapira-Lishchinsky, O., & Rosenblatt, Z. (2010b). School ethical climate and teachers' voluntary absence. *Journal of Educational Administration*, 164-181. https://doi.org/https://doi.org/10.1108/09578231011027833

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (2012). A Comparative Study of Various Tests for Normality. *Journal of the American Statistical Association*, *63*(324), 1343–1372. https://doi.org/https://doi.org/10.1080/01621459.1968.10480932

Smith, E. (2008). itfalls and promises: The use of secondary data analysis in educational research. *British Journal of Educational Studies*, *56*(3), 323-339. https://doi.org/https://doi.org/10.1111/j.1467-8527.2008.00405.x

Smith, J. A. (2015). *Qualitative Psychology: A Pratical Guide to Research Methods* (4th ed.). SAGE Publications. https://us.sagepub.com/en-us/nam/qualitative-psychology-a-practical-guide-to-research-methods/book263225

Smulders, P. G. W. (1984). Balans van 30 jaar ziekteverzuimonderzoek. *NIPG-TNO*.

Speas, C. M. (2010). Teacher absences: Types, frequency, and impact on student achievement. *Wake County Public School System*. http://www.wcpss.net/evaluation-research/reports/2010/0937teacher-absece08.pdf

Steers, R. M., & Rhodes, S. R. (1978). Major influences on employee attendance: A process model. *Journal of Applied Psychology*, *63*(4), 391-407. https://doi.org/https://doi.org/10.1037/0021-9010.63.4.391

Stewart, W. F., Ricci, J. A., Chee, E., & Morganstein, D. (2003). Lost Productive Work Time Costs From Health Conditions in the United States: Results From the American Productivity Audit. *Journal of Occupational and Environmental Medicine*, *45*(12), 1234-1246. https://doi.org/https://doi.org/10.1097/01.jom.0000099999.27348.78

Tadesse, S., Ebrahim, K., & Gizaw, Z. (2015). Sickness absenteeism and associated factors among horticulture employees in lume district, southeast Ethiopia. *Journal of Occupational Medicine and Toxicology*, *10*(1). https://doi.org/https://doi.org/10.1186/s12995-015-0074-5

Tukey, J. W. (1977). Exploratory data analysis. *Reading/Addison-Wesley*.

Usman, S., Madi, A., & Suryadarma, D. (2007). Patterns of Teacher Absence in Public Primary Schools in Indonesia. *Asia Pacific Journal of Education*, *27*(2), 207-219. https://doi.org/https://doi.org/10.1080/02188790701378826

VandenHeuvel, A., & Wooden, M. (1995). Do Explanations of Absenteeism Differ for Men and Women? *Human Relations*, *48*(11), 1309-1329. https://doi.org/https://doi.org/10.1177/001872679504801104

Varoquaux, G., & Colliot, O. (2023). Evaluating Machine Learning Models and Their Diagnostic Value. *Machine Learning for Brain Disorders*, 601-630. https://doi.org/https://doi.org/10.1007/978-1-0716-3195-9_20

Verdonk, P., Bendien, E., & Appelman, Y. (2022). Menopause and work: A narrative literature review about menopause, work and health. *WORK*, *72*(2), 483-496. https://doi.org/10.3233/WOR-205214

VfPf. (2024). *Ziekteverzuim in het onderwijs*. https://www.vfpf.nl/ziekteverzuim-in-het-onderwijs

Vijayasingham, L., & Mairami, F. F. (2018). Employment of patients with multiple sclerosis: the influence of psychosocial-structural coping and context. *8*, 15-24. https://doi.org/https://doi.org/10.2147/dnnd.s131729

Virtanen, M., Ferrie, J. E., Singh-Manoux, A., Shipley, M. J., Stansfeld, S., Marmot, M., Ahola, K., Vahtera, J., & Kivimäki, M. (2011). Long working hours and symptoms of

anxiety and depression: a 5-year follow-up of the Whitehall II study. *Psychological Medicine*, *41*(12), 2485-2494. https://doi.org/https://doi.org/10.1017/s0033291711000171
Vistnes, J. P. (1997). Gender Differences in Days Lost from Work Due to Illness. *ILR Review*, *50*(2), 304-323. https://doi.org/https://doi.org/10.1177/001979399705000207
Vuorio, T., Suominen, S., Kautiainen, H., & Korhonen, P. (2019). Determinants of sickness absence rate among Finnish municipal employees. *Scandinavian Journal of Primary Health Care*, *37*(1), 3-9. https://doi.org/10.1080/02813432.2019.1568710
Werang, B. R., Leba, S. M. R., & Pure, E. A. G. (2017). Factors influencing teacher absenteeism in the remote elementary schools of Indonesia: Empirical proof from southern Papua. *International Journal of Management in Education*, *11*(3), 223-247. https://doi.org/10.1504/IJMIE.2017.084923
Wonders, A. (2021). Reducing absence behaviour among truck drivers. *University of Twente*.
Ybema, J. F., Smulders, P. G. W., & Bongers, P. M. (2010). Antecedents and consequences of employee absenteeism: A longitudinal perspective on the role of job satisfaction and burnout. *European Journal of Work and Organizational Psychology*, *19*(1), 102-124. https://doi.org/https://doi-org.ezproxy2.utwente.nl/10.1080/13594320902793691
Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959. https://doi.org/https://doi.org/10.1093/biomet/87.4.954
Zeytinoglu, I., Lillevik, W., Seaton, M. B., & Moruz, J. (2004). Part-Time and Casual Work in Retail Trade: Stress and Other Factors Affecting the Workplace. *Industrial Relations*, *59*(3). https://doi.org/https://doi.org/10.7202/010923ar
Zuba, M., & Schneider, U. (2013). What Helps Working Informal Caregivers? The Role of Workplace Characteristics in Balancing Work and Adult-Care Responsibilities. *Journal of Family and Economic Issues*, *34*(4). https://doi.org/10.1007/s10834-012-9347-7

APPENDICES

APPENDIX A: INTERVIEW GUIDE

INTRODUCTION

- Introduce yourself and briefly share your academic background and research topic.
- Ask the participant to briefly introduce themselves.
- Request informed consent for both audio recording and analysis. Emphasize that the participant's responses will remain confidential and be processed anonymously in reports. The participant can stop at any time without consequences.

INTERVIEW QUESTIONS

Understanding Absenteeism

- How would you describe the concept of 'teacher absenteeism'?
    - And how do you experience absenteeism?
    - What is the approximate absenteeism rate at your school?
- Can you name specific circumstances or situations within the school that influence teacher absenteeism?
- In your opinion, what are the biggest differences between short-term and long-term absenteeism among teachers? What do you think influences long-term absenteeism differently than short-term absenteeism?

Personal Factors

- Are there personal factors that you think influence the decision to take leave? (age, gender, marital status)

Work-Related Factors

- Are there work-related characteristics that impact sick leave? (e.g., job satisfaction)
- What factors make you feel motivated to come to work?
    - And what factors make you less inclined to be present?
- Are there challenges in the daily work environment (e.g., classroom facilities, student behaviour) that might be less well-known but still affect teachers' decisions to come to work?
- What is the influence of other teachers on absenteeism? (absenteeism culture)

- Is it common for teachers at your school to work outside regular hours (e.g., grading, planning)?
    - To what extent does this affect absenteeism?
- What is the influence of school leadership on absenteeism among teachers?

Socio-Economic Factors

- Are there socioeconomic factors that you think influence the decision to take leave? (e.g., salary level, type of contract?)
- Are there aspects of how the school is managed or how teachers are deployed that you think contribute to absenteeism and might be unique to this school environment?

Additional

- Are there new trends or changes (e.g., in technology or policy) that you believe are influencing teacher absenteeism?
- Do you notice higher absenteeism in certain subjects (e.g., mathematics or Dutch)?
    - If so, in what way does this play a role?
- Are teachers at your school assigned to multiple locations?
    - If so, to what extent do you think having to switch between different school locations during the workweek impacts teacher absenteeism?
    - Do you see effects on attendance and workload because of this?
- Do you see a difference in absenteeism between teachers with first-degree and second-degree teaching qualifications?

CLOSING

Show the participant Table 2: list of factors that are known to influence teacher absenteeism:

- In your opinion, are there any other reasons, not typically recognized, that contribute to teacher absenteeism here?
- Is there something that you are willing to mention that we have not discussed.
- Thank participant for their time and input; participants can always contact at any point in time

DEBRIEFING

Address any questions or concerns raised by the participant during the interview.

APPENDIX B: Q-Q PLOTS AND HISTOGRAMS OF DEPENDENT VARIABLES

Absence duration before transformation

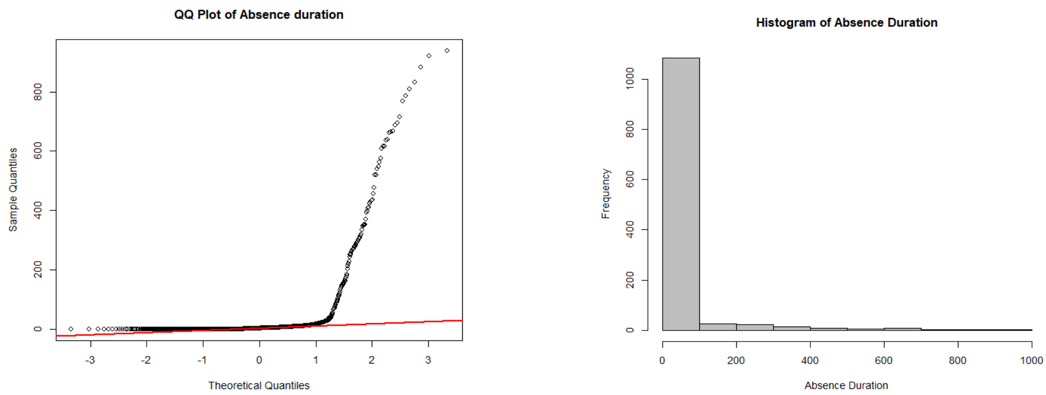

Figure 9 Q-Q Plot and Histogram of Absence Duration before Transformation

Absence duration after Box-Cox and Yeo-Johnson transformation

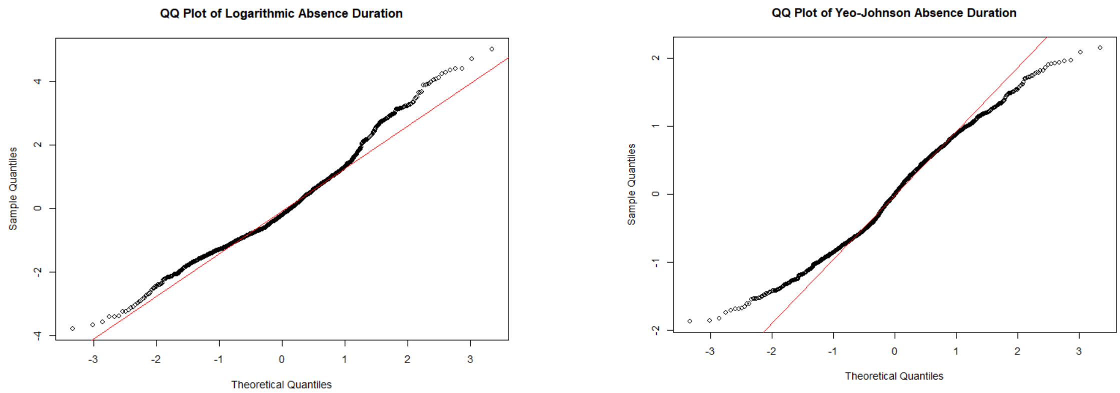

Figure 10 Q-Q Plot and Histogram of Absence Duration after Transformation

Absence frequency

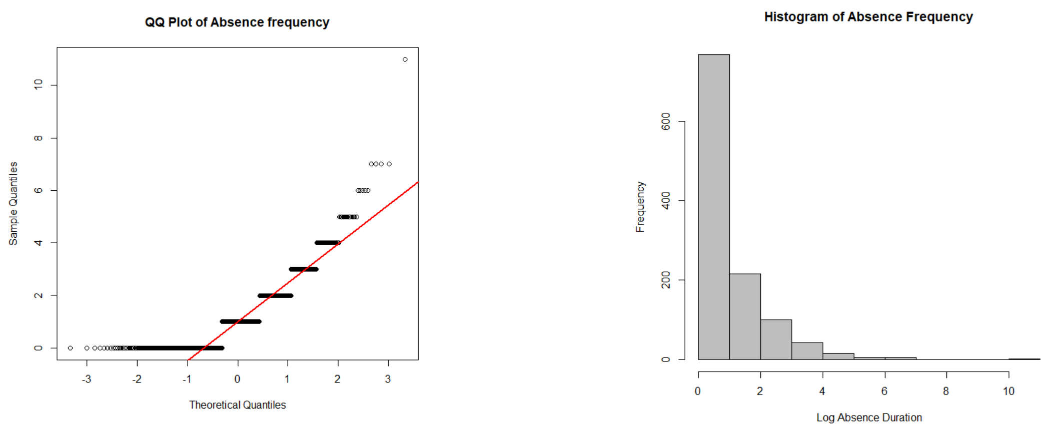

Figure 11 Q-Q Plot and Histogram of Absence Frequency

APPENDIX C: ROC CURVES

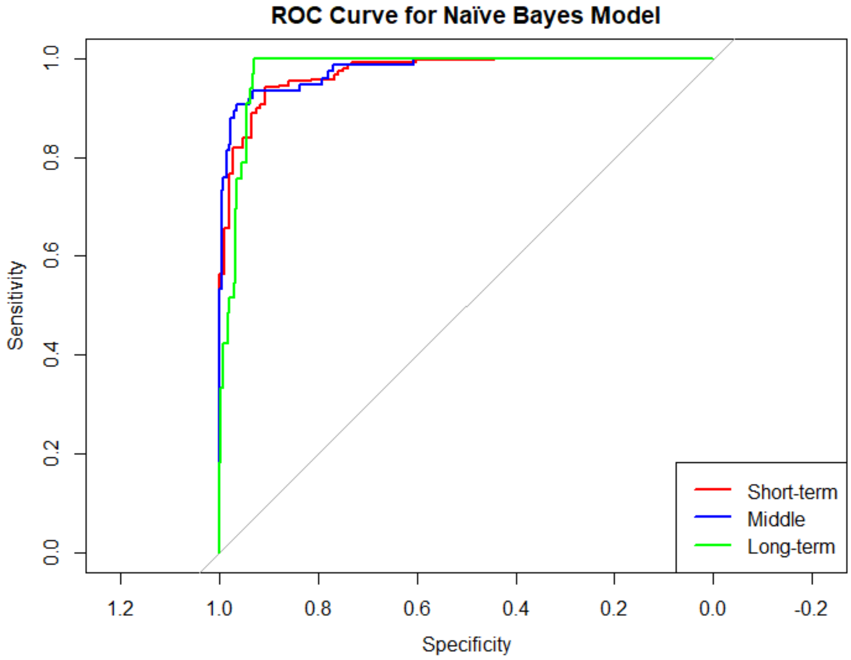

Figure 13 ROC Curve - Naïve Bayes

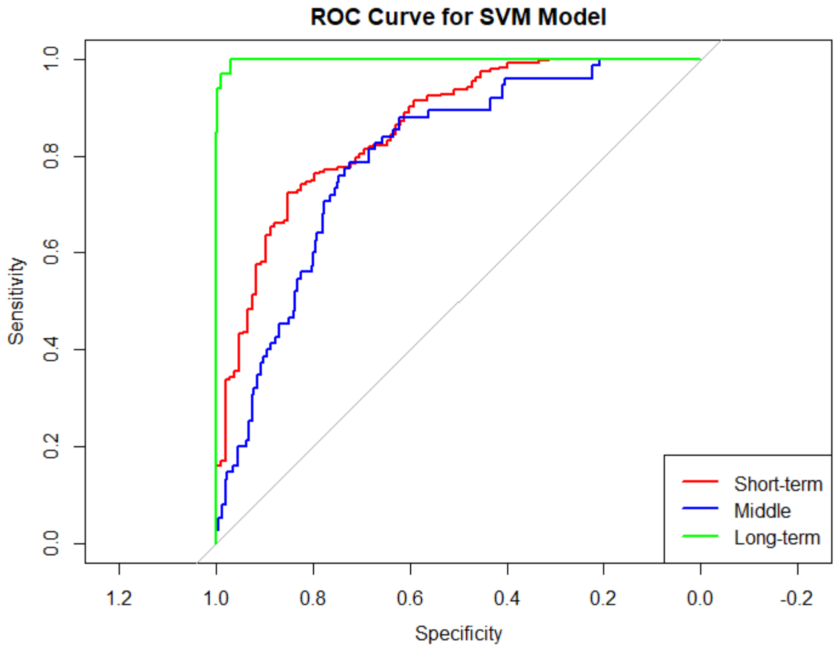

Figure 12 ROC Curve - SVM