Examining the Impact of Data-Informed Decision-Making Approaches on Goal Setting and Action Planning Among High School Teachers

Steven Bos

Faculty of Behavioral, Management and Social Sciences, University of Twente

Educational science and technology,

1st Supervisor: Prof. Dr. Kim Schildkamp

2nd Supervisor: Lucas Silva

March 21, 2024

Abstract

Data's growing influence on decision-making extends to education in the form of Data-Informed Decision Making (DIDM). This study investigated how DIDM approaches which start with either data or goal setting, influence goals and action plans. Furthermore, this study explores the influence of teacher efficacy, a teacher's beliefs in their capacity to effectively educate, on goals and goal setting.

High school mathematics teachers were randomly assigned to the goal-first or data-first groups. The goal-first group sets an end-of-year goal grade and administers a test to assess the current level of students, after which an action plan is formulated. The data-first group administers the test, sets the goal and lastly formulates an action plan. Post-experiment interviews explore the goals, action plans, and efficacy beliefs.

Using a mixed linear model a non-significant positive relationship between a goal-first approach and goal differentials, defined as the numerical difference between the current performance level and the goal, was established at a P-value of .108. Regarding action plans, the goal-first group focused more on long-term comprehension by adding lessons, focusing on study habits, and repeating the materials. The data first group focused on attaining the goal efficiently, using homework and buffer lessons where possible.

Keywords: Data-Informed Decision Making, Action Plan, Goal-Setting, Teacher Efficacy.

Contents

Abstract 1
Contents 2
Introduction 3
Theoretical Framework 6
Data-Informed Decision Making 6
Goal-First DIDM Approach 8
Data-First DIDM Approach 9
Goals and Achievement 11
Action Plans 12
Teacher Efficacy 13
Method 14
Research Design 14
Participants 15
Instrumentation 16
Procedure 18
Data Analysis 19
Results: 22
Quantitative Results 22
Qualitative Results 25
Discussion 37
Main Findings 37
Recommendations 40
Limitations 41
Future Research 43
Conclusion 43
Bibliography 45
Appendix I: Answer key test 3 (Dutch) 50
Appendix II: Interview Schema (Dutch) 53
Appendix III: Plots 56
Appendix IV: R-code 58
Appendix V: Use of Generative Models 60
Appendix VI: Codebook 61

Introduction

Data is gaining prominence as a key influencer of decision-making across a wide range of industries, including education. As technology continues to evolve, the amount of available data is increasing at an unprecedented rate. Organizations have the opportunity to leverage this increase to gain valuable insights and make informed decisions (Sarker, 2021). The education sector has steadily embraced the use of data to shape policy-making, curriculum design, and teaching methods (Campbell & Levin, 2009). Today's educators increasingly leverage data to pinpoint student challenges, track progress, and personalize learning for individual needs (Baig et al., 2020).

Data use in education can positively affect educational outcomes as has been shown in literature reviews by Marsh (2012) and Grabarek & Kallemeyn (2020). An example of a study highlighting this is van Geel et al. (2016). This study compared two groups of schools: a control group and an experimental group that received training in effective data-informed decision-making (DIDM), including data analysis, instructional planning, and evaluation. After two years, the experimental group showed educational improvements equivalent to an additional month of schooling, measured in mean mathematical ability score.

However, while van Geel et al. (2016) highlight the positive impact of data-use interventions, they also underscore the need for comprehensive research that simultaneously examines both the outcomes and processes of such interventions. DIDM can start with either data or goal setting. The goal-first approach starts with goal-setting and subsequently employs data to investigate why goals (have not) been reached and formulate action plans accordingly (Schildkamp, 2019). The data-first approach starts by collecting and analyzing data to identify areas for improvement and establishing goals and action plans accordingly (van Geel et al., 2016).

This study aims to determine whether there is an optimal starting point for data use approaches in education, focussing on the effects of beginning with either data collection or goal setting. Examining this can aid researchers in the development and improvement of DIDM approaches and support educators and stakeholders in determining which DIDM approach best suits their unique circumstances.

Aligning with the need for research put forth by van Geel et al. (2016), this study aims to identify whether there is an optimal approach by comparing the effects of both approaches on goal setting and action planning. To investigate this the relative height of goals, referred to as goal differentials, and the corresponding action plans to achieve these goals are used.

Locke & Latham (2019) suggest, that setting higher, yet realistic goals tends to produce better results. For this study, we will use the term ‘goal differential’. The term 'goal differential' refers to the numerical difference between the current performance level and the goal. In this context, the goal is defined as the desired classroom average score by the end of the academic year. Goal differentials are deemed an appropriate metric to estimate future educational improvement due to their strong positive relationship with educational results. Hence, the first research question (RQ1) was formulated: “*What is the difference in goal differentials between teachers using a goal-first and a data-first DIDM approach?”*

The path from goals to results is described in action plans. Action plans are systematic approaches to achieving the set goal. The analysis of action plans aims to investigate whether the goal-first and data-first approaches lead to variations in educational approaches concerning the attainment of the goal grade. To this end RQ2 was formulated: *“How do teachers using goal-first and data-first DIDM approaches differ in their goal setting and formulation of action plans”.*

Keuning et al. (2017) conducted a comparative analysis among schools to assess the effectiveness of Data-informed Decision Making (DIDM) interventions. Their findings

revealed that teachers with positive attitudes toward data and higher literacy skills were more likely to achieve favourable outcomes with DIDM interventions. Likely these educators were more inclined and better equipped to use data for enhancing educational practices. Dunn et al. (2013) demonstrated that teacher efficacy is positively correlated with positive attitudes toward data and the willingness to work with data and to develop data literacy skills. By integrating the findings of these two studies, it becomes apparent that teachers with high efficacy potentially possess characteristics which aid the implementation of DIDM approaches. Scheer and Visscher (2016) indicate a reciprocal relationship between efficacy and DIDM, high efficacy fosters receptiveness to DIDM methods and indicates the necessary reflective capabilities for effective use of DIDM. Additionally, their study shows that sustained practice with a DIDM approach leads to increases in teacher efficacy over time.

Teacher efficacy can be defined as a teacher's perception of their teaching abilities, high efficacy teachers see themselves as capable teachers who can make an impact on the learning trajectory of students (Timperley & Phillips, 2003). Teachers who possess a high teacher efficacy, tend to establish more ambitious goals (Ross & Bruce, 2007). Consequently, teachers with high efficacy scores tend to establish challenging and purposeful goals, fostering a more productive learning environment (Tschannen-Moran et al., 1998). The role of teacher efficacy in goal setting and achievement (Locke & Latham, 1990) and in DIDM attitudes and adoption (Dunn et al., 2013) prompted the formulation of RQ 3 "*How does teacher efficacy influence goal differentials and respective action plans?"*

Together the three research questions encompass a comprehensive examination of goal differentials and action plans within the framework of DIDM and teacher efficacy. Based on this a comprehensive understanding of the effective use of data in educational settings can be constructed, leading to the identification of the most optimal DIDM approach for educators.

Theoretical Framework

Data-Informed Decision Making

Data-informed decision-making (DIDM) is a systematic approach that involves collecting and analysing data to enhance student and school performance (Ikemoto & Marsh, 2007; van Geel et al., 2016). This process has been conceptualized in various ways within the academic literature.

Schildkamp (2019) identifies four key elements of DIDM geared towards educational improvement: goal setting, data collection, sense-making, and action and evaluation. Similarly, van Geel et al. (2016) propose a model encompassing the evaluation and analysis of results, setting SMART and challenging goals, devising strategies for goal achievement, and implementing these strategies. Lai & Schildkamp (2013) outline a five-step process: defining the purpose, collecting data, analysing it, interpreting results, and taking action. Mandinach (2012) states that DIDM has 4 critical steps, data collection, analysis and interpretation of data, instructional decision-making based on the findings, and finally a positive effect on learning outcomes. These models highlight an ongoing debate among scholars about the optimal structure and sequence of DIDM steps, particularly whether goal setting or data collection should precede the other (Schildkamp, 2019; Van Geel et al., 2016).

Despite differences, common elements emerge across these models, such as defining purpose, collecting, analysing, and interpreting data, setting goals, strategizing, executing plans, and evaluating outcomes. While not every model explicitly includes each of these stages, collectively they offer a comprehensive understanding of DIDM.

This paper focuses on the starting point of DIDM processes; the step of defining purpose is seen as a prerequisite and is not viewed as a starting point. Hence, the DIDM models start with either goal-setting (Schildkamp, 2019) or data collection (Van Geel et al., 2016). The subsequent phases of the models generally follow a more agreed-upon structure,

although slight variations in phases can occur depending on whether goals are already established.

Defining the purpose of DIDM involves identifying the reasons for employing DIDM and determining the necessary data types (Lai & Schildkamp, 2013). This initial stage, while not always explicitly stated, as seen in van Geel et al.'s (2016) data-first approach, is inherently present. A purpose may range from monitoring educational progress to addressing specific research questions, such as investigating the effectiveness of a new teaching method.

Goals, on the other hand, are structured and measurable objectives. The goal-setting stage in these models emphasizes the importance of establishing clear, structured goals to guide the data-driven improvement of educational outcomes. SMART goals are particularly advocated (van Geel et al., 2016; Schildkamp, 2019).

Data collection can involve several types of data, however, within the context of this study only formal and informal data are relevant. Formal data comprises systematically gathered qualitative and quantitative information about students, parents, teachers, and stakeholders. Informal data refers to information collected by teachers through everyday interactions with students, such as observation and conversation (Schildkamp, 2019).

Data analysis and interpretation are critical, here the data is sorted, analysed, and transformed into actionable insights, leveraging experience, expertise, and understanding of the educational context (Schildkamp, 2019; Lai & Schildkamp, 2013). These insights can inform goal-setting and action planning. In a goals-first approach, however, goals are set earlier and drive data collection and subsequent phases.

Poortman and Schildkamp (2016) identified three key components of action plans: curriculum, assessment, and instruction all of which can be implemented in a variety of ways. For instance, enhancing the curriculum could involve improving its coherence by illustrating the relationships between various mathematical topics. Assessment and instruction can serve

various purposes, including differentiation among students. Assessments help identify at-risk students, who can then receive additional instruction.

Executing and formulating an action plan in education are complex and multifaceted tasks. The implementation phase often presents unforeseen challenges. These obstacles can emerge as discrepancies between theoretical plans and real-world scenarios, requiring educators to adapt and respond to situations not initially anticipated (Schildkamp, 2019).

Finally, an evaluation stage assesses the effectiveness of the DIDM process by determining whether the goals set earlier have been reached through the actions taken (Schildkamp, 2019; Ikemoto & Marsh, 2007).

Goal-First DIDM Approach

In the goals-first methodology as outlined by Schildkamp (2019), a broad purpose informs the setting of SMART goals. The SMART goals drive the collection of data. Hereafter a thorough analysis and interpretation of the data are crucial, serving not only to assess the current level of student achievement in relation to the set goals but also to identify areas for strategic intervention based on the data (i.e., causes for not reaching a certain goal). Insights gained from this process inform the formulation of a targeted action plan.

In this approach, teacher intuition is considered the guiding factor for setting goals, with the belief that educators, through their experience and insights, can make precise predictions about the current and future abilities of students. Baron-Thiene & Alfermann (2015) showed that teachers could predict student achievement with reasonable accuracy. Professional development and the frequency of formative assessment had a positive correlation with the accuracy of intuition. Another study into teacher intuition showed teachers performed comparable to early warning systems when estimating college enrolments indicating the teachers had a good understanding of the academic potential of students (Soland, 2013).

However, human judgment and intuition are fallible. The accuracy of intuitive judgment is closely tied to the familiarity and complexity of a situation. In situations characterized by greater familiarity and lower complexity, judgments tend to be more precise (Klein, 2017; Betsch, 2008). Wang et al (2018) showed teacher expectations can be affected by nonrelated student characteristics such as ethnicity, gender, socioeconomic status, and attractiveness. There is also the risk of confirmation bias (Vanlommel et al., 2017), which occurs when people prioritize information that confirms existing beliefs over other information (Kahneman & Frederick, 2005). In their study, Hollenstein et al. (2023) illustrate how low expectations from teachers can negatively affect student achievement, and self-belief, and increase anxiety related to mathematics. However, it is important to acknowledge that positive expectations have the opposite effect. Therefore, biased judgments stemming from poor intuitive assessments can result in both advantageous and disadvantageous outcomes for student. Rubie-Davies et al. (2006) illustrate negative expectation effects using the Māori within the New Zealand school system, showing their relative academic improvements decline over time due to negative expectations.

Lastly, the goal-first approach carries the risk of teachers setting target grades below current ability levels. While such instances may be rare, they are possible; however, no empirical data on this point is currently available. Teachers' intuitions are fallible and current ability levels can be misjudged in the absence of data. In such cases, a transition towards a data-first strategy becomes necessary to establish new, more appropriate goals.

Data-First DIDM Approach

In the data-first approach, a broad purpose such as improving mathematics grades is formulated. Based on this purpose, data are collected. Through analysis, this data is transformed into information, which, when combined with stakeholder understanding, generates actionable knowledge (Ikemoto & Marsh, 2007). This knowledge identifies

potential areas for improvement and enables the formulation of SMART goals and accompanying action plans (van Geel et al., 2016).

Data are viewed as the logical starting point as it highlights areas requiring improvement, which then informs the establishment of goals aimed at addressing these gaps. By analysing data and identifying a learner's current strengths and weaknesses, educators can effectively pinpoint tasks and challenges that fall within the learner's Zone of Proximal Development (ZPD), ensuring that the learning experience is appropriately tailored to their developmental level, maximizing their cognitive growth (Shabani et al., 2010). A quote by a school administrator from an interview study by Earl and Fullan (2003, p. 390) illustrates this: "We looked at the data, identified the weaknesses, and started to plan improvements. The school started to make immense progress in teaching and learning because we could see the problem and we got the support for the teachers."

However, Schildkamp and Datnow (2022) hypothesize that a data-first approach might lead to less ambitious goal-setting and could even risk schools becoming complacent with moderate outcomes, potentially dissuading efforts to achieve higher standards. The sheer volume of data available in schools today adds to the complexity, potentially overwhelming teachers and averting the focus from other critical aspects of data-informed decision-making (Krein & Schiefner-Rohs, 2021). Teachers often hold negative attitudes toward data use and lack data literacy skills (Jimerson, 2014; Mandinach & Gummer 2013). This, in turn, can lead to low teacher motivation and an inability to determine the quality and meaning of data (Schildkamp & Datnow, 2022).

Additionally, limited expertise and knowledge make teachers more vulnerable to cognitive biases. One such bias is the anchoring bias: when individuals are presented with a singular piece of information, such as results on a single test, they tend to overly rely on this piece of information when forming specific judgments (Tversky & Kahneman, 1974).

In conclusion, while the data-first approach offers a structured pathway towards educational improvement, it also brings inherent risks and challenges. Teacher resistance, teacher anchoring bias and the overwhelming volume of data are significant hurdles. Additionally, it might lead to less ambitious goals which can have a problematic impact on academic achievement in the future.

Goals and Achievement

Setting goals is an effective strategy to drive achievement when certain conditions are met (Locke & Latham, 2006). Utilizing SMART goals provides a focused and achievable approach, leading to measurable outcomes such as improved standards, enhanced student progress, and a higher quality of education (Carvalho et al., 2022).

When individuals, in this case, teachers and students, are dedicated to their goals, possess the necessary abilities, and are not hindered by conflicting aims, there is a strong and direct relationship between the challenge of the goal and their performance. Demanding goals elicit greater effort and persistence compared to moderate, easy, or vaguely defined goals (Locke & Latham, 2019).

However, it is important to consider goals set in the distant future may lose their motivational value, as they can feel disconnected from immediate actions and incentives (Bandura, 1988). Therefore, assessing the relative height of goals within a reasonable timeframe offers a practical indicator of future achievement. To this end the term goal differential has been introduced in this paper, a goal differential is the distance between the current ability and the desired ability expressed by the goal.

High but realistic goals provide a valuable foundation for improvement (Lock & Latham, 2019). However, only goals are not enough to enable improvement, goals should be combined with relevant data to formulate a comprehensive action plan.

Action Plans

An action plan is a systematic approach to achieving a goal and determining how to assess whether the goal has been met. In the literature terms such as strategy (van Geel et al., 2016) or improvement plan (Carvalho et al., 2022) are used as synonymous for action plan. Action plans can encompass all levels of the school hierarchy, they can be school-wide or focused on a single student. In the context of this paper, the focus lies on action plans created by teachers focused on improvement at the classroom level.

To ensure clarity during implementation, planned actions must be clear, specific, and well-defined. Action plans should detail practical steps, offering a roadmap for turning intentions into actions (Carvalho et al., 2022). Poortman and Schildkamp (2016) show that these action plans often include one or more of the following aspects: improving curriculum coherence, developing, and implementing assessment, and providing additional instruction. Assessment and instruction frequently complement each other in differentiation. Assessments pinpoint struggling students, who then receive supplementary instruction to improve their performance.

The quality and the effectiveness of an action plan are dependent on many variables. Carvalho et al (2022) identified criteria based on which the quality of an action plan is measured. Firstly, the action plan should align with the priorities and context of the school and its student population. Second, the plan should be comprehensive linking needs, goals, and solutions. Furthermore, the plan should have strong foundations. Evidence-based, research-based, and data-based strategies are essential components, underlining the importance of grounded practices. Additionally, parental and community involvement can add a valuable dimension, recognizing the collaborative effort needed for success (Carvalho et al., 2022).

Teachers play a pivotal role in both developing and implementing action plans. Their characteristics, especially teacher efficacy, significantly influence this process. Research indicates that teachers with a higher sense of efficacy are more inclined to employ formative

assessment methods (Allinder, 1995) and a wider array of instructional strategies (Opper, 2019). This highlights the importance of considering teacher efficacy.

Teacher Efficacy

Teacher efficacy refers to teachers' confidence in their ability to effectively educate and guide students (Timperley & Phillips, 2003). It is based on self-efficacy, which can be described as individuals' beliefs about their capacity to perform behaviours necessary to achieve specific goals (Bandura, 1977). Tschannen-Moran and Hoy (2001) identified three dimensions of teacher efficacy: instructional strategies, classroom management, and student engagement. Efficacy for Instructional Strategies refers to the teacher's self-perceived proficiency in employing diverse instructional techniques, including using various assessment strategies, providing alternative explanations, and crafting effective questions (Tschannen-Moran & Hoy, 2001). Efficacy for Classroom Management relates to the teacher's self-perceived ability to maintain a well-managed classroom environment, involving controlling disruptive behaviour, setting clear behaviour expectations, and establishing effective routines for classroom activities. Efficacy for Student Engagement focuses on the teacher's self-perceived effectiveness in engaging students, fostering a positive learning environment, and promoting student motivation and interest in learning. The dimensions of teacher efficacy underscore the depth of teachers' professional experiences and the essential elements of effective teaching (Tschannen-Moran & Hoy, 2001).

Teacher efficacy plays a vital role in goal setting, response to feedback, commitment, and adoption of effective strategies (Locke & Latham, 1990). It influences teachers' perception of their competence and the potential impact they can have on students' learning outcomes. Efficacy beliefs act as self-fulfilling prophecies, shaping behaviour control and mental resilience in the face of failure (Bandura, 1982). Higher efficacy leads to greater perceived control and resilience.

Teachers exhibiting high levels of efficacy, as determined by the average scores across the three previously mentioned dimensions, tend to establish more ambitious goals for both them and their students (Ross & Bruce, 2007). They demonstrate increased effort, enthusiasm, organization, clarity, and a willingness to explore innovative and challenging teaching ideas and techniques (Allinder, 1994; Ross, 1998). Consequently, teacher efficacy may significantly influence goal differentials and action plan formulation.

Method

This chapter discusses the various aspects of the research method and design. The research design will be discussed, followed by a description of participants, instrumentation, and data analysis.

Research Design

To investigate the effects of DIDM approaches and teacher efficacy on goal-setting and action plans, both qualitative and quantitative methods were used. The use of both types of methods can provide a deepened understanding (Shorten & Smith, 2017). The design consisted of a two-group post-test-only experiment followed by interviews. Participants were randomly assigned to either the goal-first group or the data-first group.

In the two-group post-test-only experiment the independent variable was the DIDM approach, and the dependent variable was the goal differential. In the goal-first group, participants started by examining the test and setting a goal grade for the end of the school year per student. Subsequently, the students took the test, the participant graded, analyzed, and interpreted the results, and formulated an action plan. In the data-first group, participants started by administering the test, followed by analyzing and interpreting the results, after which the goal grades were set, and an action plan was formulated.

To align with real-life teaching scenarios, participants had flexibility in developing their action plans based on their individual needs. The plans did not have to follow a

mandated format and were allowed to be documented formally or outlined mentally. The core requirement was for teachers to engage in a thoughtful process, envisioning practical methods and approaches to achieve their set goals.

Following this stage, interviews were conducted with all available participants. These interviews focussed on how the goals were set, the development of action plans, the content of the action plans and finally the efficacy beliefs of teachers.

Participants

The sample of this study consists of high school teachers instructing mathematics for grades 2 and 3 in HAVO/VWO, a level of secondary education designed for students aged 12-14, preparing them for higher education or university. The sample composition consisted of nine participants, four male and five female teachers, with a slightly unbalanced representation of three from grade 2 and six from grade 3. The study involved participants with varying levels of experience ranging from under five years to over twenty years. The participants worked at eight schools spread across the Netherlands. This ensured a sample from diverse geographical and educational backgrounds.

The sampling strategy employed was convenience sampling. One participant was identified through the supervisor's professional network, while the remaining eight were recruited via an extensive email campaign targeting high schools across the Netherlands. The campaign utilized various channels, including existing newsletters, direct emails to school faculty, general information emails (@info), and personalized outreach to mathematics teachers. Prospective participants were invited through a digital flyer which described the study requirements and procedure.

Instrumentation

The Math Test

The mathematics assessment originates from a study by de Vries et al. (2022) focusing on assessing the general mathematics level of 12-14-year-old students in the first three grades of senior general and pre-university secondary education. The questions for the four mathematical exams were formulated based on the learning targets of these grades in the countries involved in the Erasmus+ Project. Each exam contains 13 to 30 open-ended problems. Four versions of the mathematics test are available, specifically designed for each year within the first three years of HAVO/VWO, including an additional version for the post-evaluation of third-year students. The test will assess mathematical abilities by covering a range of topics among which are number pattern recognition, geometry, arithmetic, and percentages.

For this study, only the tests for the second and third grades were utilized, which contain 28 and 13 problems respectively. In the pretest phase of de Vries et al.'s (2022) study, a Rasch analysis was conducted. The obtained reliability coefficients between items were notably high, with values of 0.98 for grade two and 0.96 for grade three. These results indicate a substantial level of consistency in ranking and differentiating the difficulty levels of items on both tests. Conversely, the reliability coefficients between persons varied between the two tests. For the grade two test, a coefficient of 0.79 suggested a relatively high level of consistency in ranking individuals based on their abilities. However, the grade three test exhibited a lower reliability coefficient of 0.46, indicating a decreased level of consistency in ranking individuals. This indicates the test has a limited ability to differentiate between students of varying ability levels.

Tests 2 and 3 were adapted to suit the specific requirements of the research. The tests originally contained a metacognitive element, which was omitted for this study. Additionally, the amount of personal information required from students was minimized, as the tests were

administered directly by the teacher instead of the researcher. Lastly, the answer keys were streamlined to facilitate ease of use by clarifying the scoring process and eliminating instructions that were relevant only to researchers. For additional context, Appendix I contains the modified answer key for test 3, this key includes the complete set of questions, answers, and the scoring method.

Interview Schema

Regarding the interview, a semi-structured approach was used to explore the teachers' goal-setting, action plans and teacher efficacy. This method strikes a balance between structured questioning and open-ended discussions, allowing for comparative analysis while also providing the flexibility to delve into unforeseen aspects (Babbie, 2021). The interviews explored the rationale behind goal setting, the formulation and structure of action plans, and teacher efficacy.

The interview schema for this study was designed based on the literature findings discussed in the previous chapter. For goals setting questions pertaining to, realism, ambitiousness, motivation, and alignment with school plans were included based on Lock and Latham (2019), Bandura (1988) and Carvalho et al (2022). Regarding action plans, questions covering the content of the curriculum, assessment, differentiation, evidence-based practices, and parental involvement were added based on Poortman and Schildkamp (2016), Sadler (1989), and Carvalho et al. (2022). Finally, teacher efficacy questions followed Tschannen-Moran and Hoy's framework (2001), exploring efficacy in instructional strategies, classroom management, and student engagement. The Dutch interview schema can be found in Appendix II.

Procedure

Math Test

The mathematics test was administered by teachers in a regular classroom test setting, where students sat separately without communication and only had access to materials relevant to the test (e.g., pen, paper, question sheet) (De Vries et al., 2022). The test took approximately 40 minutes and was scored by the teacher based on the provided answer keys. The answer key employed a strict scoring system, adhering closely to a binary format of right or wrong. This contrasts with traditional practices in Dutch education when teaching mathematics, where partial credit is often awarded for partial solutions or using correct methods. The study's approach facilitates uniformity in grading.

Interview

The interviews were conducted online using Microsoft Teams. The interviews occurred within two weeks of the teachers finishing the experimental phase, to ensure their reasoning and thought processes were still salient. Participants received an email before the interview, outlining the process. The email provided an overview of the interview protocol, a copy of the interview schema for reference, and information on how the interview data would be used, ensuring transparency in the research process. The interviews lasted approximately 40 minutes. At the start of each interview, consent was obtained for recording the session. The recording was automatically transcribed using the embedded transcription function from Microsoft Teams. The transcriptions were checked for errors and anonymized after which the recording was deleted. Open-ended questions were used to encourage detailed answers, with the interviewer having the flexibility to ask follow-up questions for clarification or deeper exploration of specific topics.

Data Analysis

Quantitative Analysis

Quantitative analysis was conducted using R-studio. The variables of interest were the dependent variable goal differentials and the independent variable DIDM approach. Goal differentials were calculated by subtracting the test scores from the goals, indicating the desired improvement in classroom performance. The DIDM approach is a dichotomous variable indicating whether participants belong to the data-first or goal-first group.

To answer RQ1 "*What is the difference in goal differentials between teachers using a goal-first and a data-first DIDM approach?"* Two-sample t-tests were conducted to compare the mean goal differential scores between teachers (N = 9) using different teaching approaches (goal-first vs. data-first) and between students (N= 198) taught by teachers with different approaches.

To assess the normality of both groups, a Shapiro-Wilk test was conducted. The results indicated that the goal differentials in both the data-first and goal-first groups did not significantly deviate from normality, with *p-values* of .153 and .492, respectively, supporting the null hypothesis that the data is drawn from a normally distributed population. Therefore, employing a normal t-test for subsequent analyses was deemed appropriate.

Given the inconclusive results of the two T-tests, where one indicates the significance and the other does not, a mixed linear model was constructed to further assess the impact of teacher approaches on goal differentials while accounting for both fixed and random effects (van den Berg, 2021).

The model included the DIDM Approach variable as the independent variable and Goal Differential on the student level as the dependent variable. Teacher-specific effects, namely variation introduced by individual teachers, were accounted for in the random effects of the model, acknowledging the hierarchical structure of the data (van den Berg, 2021),

where teachers operate within the group variable. The dataset comprised 198 observations grouped within 9 distinct teachers.

The quality of the model was assessed using a range of statistical tools. The model has a REML criterion of 521.3, indicating a satisfactory fit to the data. Diagnostic plots, Q-Q plots and residual plots were employed to affirm the overall goodness of fit, these can be found in appendix III. Examination of scaled residuals demonstrated a symmetrical distribution around zero, with most residuals falling within an interquartile range of -0.56 to 0.53. Residual plots were utilized to evaluate homoscedasticity, ensuring consistent variance of residuals across predictor variables. These observations and metrics indicate that the model fits the data adequately and meets the expectations set by the model assumptions.

Qualitative Analysis

The second research question "*How do teachers using goal-first and data-first DIDM approaches differ in their goal setting and formulation of action plans"* was answered using a qualitative analysis which involved conducting interviews, recording them, and transcribing the recordings. A hybrid coding schema combining inductive and deductive coding was used. Deductive coding was informed by existing literature on action plans. Based on the literature and the interview schema the general topics were outlined such as goals and action plan content. Subsequently, underlying themes were formulated. In the case of action plan content assessment, curriculum and degree differentiation are examples of themes. Within the theme, codes were formulated. For example, the theme of goal attitudes uses the codes realistic, ambitious, and motivating. This base was supplemented with inductive coding to capture new insights and non-standard findings emerging from the interviews (Babbie, 2021). This approach to coding aims to incorporate both expected outcomes from prior literature and unexpected observations from the data.

After the completion of the coding phase, the subsequent step involved data analysis. This encompassed scrutinizing the coded data to identify recurring patterns, prevalent themes,

and noteworthy or distinctive elements, with a particular focus on their impact on the formulation of goals and action plans. These findings were used to form a comprehensive understanding of the processes and actions of both groups. These were compared to gain insight into the similarities and disparities in action plan formulation between the data-first and goal-first groups.

The third research question "*How do teachers using goal-first and data-first DIDM approaches differ in their goal setting and formulation of action plans"* was also answered using the data from RQ2, in combination with interview data about teacher efficacy beliefs. The data about efficacy beliefs was coded using themes the dimensions outlined by Tschannen-Moran and Hoy (2001) as themes. Utterances were subsequently coded based on the degree of efficacy they expressed, with codes based on the Teacher Efficacy Scale (Tschannen-Moran & Hoy, 2001). This coding approach allowed for a comparison of action plans and goal differentials among teachers expressing different levels of efficacy. Based on this the impact of efficacy on these factors could be assessed.

To assess the reliability of the quantitative part of the study, Cohen's kappa was calculated for qualitative content analysis, ensuring the reliability of coding (Burla et al., 2008). The inter-rater reliability was checked using 12.5% of the total interviews (1 interview), yielding a Cohen's Kappa of 0.82.

Results:

Quantitative Results

In the quantitative analysis the connection between DIDM approaches and goal-differentials was explored to answer RQ1: *"What is the difference in goal differentials between teachers using a goal-first and a data-first DIDM approach?"*

Based on RQ1 H0 and H1 can be formulated:

- Null Hypothesis (H0): *There is no statistically significant difference in the goal differentials set by teachers using a goal-first and a data-first DIDM approach* (μ_goal-first - μ_data-first = 0).

-Alternative Hypothesis (H1): *There is a statistically significant difference in the goal differentials set by teachers using a goal-first and a data-first DIDM approach* (μ_goal-first - μ_data-first $\neq$ 0)

Overview

Five participants joined the data-first group, and four participants joined the goal-first group. Each tabulated the grades and goal grades for an average of 22 students, totalling 198 students. Tables 1 and 2 showcase class averages, including test scores, goal grades, and goal differentials for participants in the data-first group and goal-first group, each participant is referred to by a letter ranging from A to I.

Table 1

Class averages of data-first participants

Participant	**A**	**B**	**C**	**D**	**E**
Test score (class average)	5.38	5.30	4.09	5.28	4.04
Goal grade (class average)	6.98	7.20	5.53	7.31	7.31
Goal differential (class average)	1.60	1.90	1.44	2.03	3.27

Table 2

Class averages of goal-first participants

Participant	**F**	**G**	**H**	**I**
Test score (class average)	3.74	2.85	4.09	4.76
Goal grade (class average)	7.44	6.14	5.53	6.91
Goal differential (class average)	3.70	3.29	2.38	2.15

Descriptive Statistics

To gain an initial understanding of the central tendencies and variabilities within each group the respective descriptive statistics were calculated and analyzed, note that these statistics pertain to the goal-differentials of both groups. The mean score (*M*) for the data-first group was found to be 2.05 with a standard deviation (*SD*) of 0.74. In contrast, the goal-first group exhibited a higher mean score (*M*) of 2.88, with a standard deviation (*SD*) of 0.75. Table 3 provides an overview of the descriptive statistics of both groups.

Table 3

Descriptive statistics

	Data-first	**Goal-first**
Mean	2.05	2.88
SD	0.74	0.75
Minimum	1.60	2.15
Maximum	3.27	3.70

Inferential Statistics

A Two-sample t-test was conducted to compare the mean goal differential scores of the two groups of teachers. The results of the analysis produced a t-value of -1.70 with degrees of freedom (*df*) approximated to 6.51. The 95% confidence interval for the difference in means ranged from -2.01 to 0.34, and this interval includes zero. The corresponding p-value was

calculated to be .136. The obtained p-value (p = .136) exceeds the required alpha level of .05, which is a standard norm used in social sciences (Babbie, 2021).

A second two-sample T-test could be conducted at the student level, to see whether the goal differentials differ at the student level based on whether the teacher used a goal-first (n = 96) or data-first (n =102) approach. The T-test was conducted to compare the goal differential scores of the two groups of students. The test produced a t-value of -5.83 at 181.09 degrees of freedom (*df*). The 95% confidence interval for the difference in means ranged from -1.71 to -0.58, which does not include 0. The corresponding p-value was calculated to $p < .001$, falling within the alpha range of .05. Indicating students with a teacher using the goal-first approach attained higher goal differentials.

Mixed Model

The discrepancies in the results of the two-sample t-tests at different levels of analysis highlight the importance of considering both group-level and individual-level variations. Building a mixed linear model allows us to account for these variations and assess the impact of teacher approaches on goal differentials more comprehensively, encompassing both fixed and random effects to address inherent variability in the dataset. The fixed effects stemmed from the Teacher approach variable, capturing systematic disparities between data-first and goal-first approaches. Conversely, random effects accounted for the variations introduced by the individual teachers. The dataset comprised 198 students grouped within 9 distinct teachers. The limited number of teachers potentially limits the statistical power at the teacher level. Conducting a low-power analysis heightens the likelihood of encountering both false positives and false negatives, diminishing the reliability of both statistically significant and insignificant findings (Button et al., 2013).

Regarding random effects, variability attributed to teacher-specific effects was estimated to have a variance of 0.72, while residual variability was estimated at 0.51. This results in an ICC of 0.59. In terms of fixed effects, the impact of the second teacher approach

variable (representing the goal-first teaching approach) on goal differentials was found not to be statistically significant (*Estimate* = 0.91, *SE* = 0.49, *t-value* = 1.84, *p* = .108). This suggests no significant difference in goal differentials between the two teaching approaches after accounting for teacher-level variability. Therefore, the null hypothesis (H0): *There is no statistically significant difference in the goal differentials produced by teachers using a goal-first and a data-first DIDM approach (μ_goal-first - μ_data-first = 0),* can not be rejected.

Qualitative Results

Based on the interview data the relationship between DIDM approaches and action plans is explored first. The second part pertains to the relationship between teacher efficacy and action plans.

DIDM Approaches and Action Plans

This part aims to explore RQ2, *"How do teachers using goal-first and data-first DIDM approaches differ in their goal setting and formulation of action plans*" The analysis compares patterns and planned actions observed in interview data from both groups. Starting with goal setting followed by an action plan.

Goal setting

Data-first Group. The data-first group experienced the study procedure for goal setting as relatively similar to their normal practices. Some teachers did indicate setting individual goals at the test level for all students was more comprehensive than their usual approach. While these teachers did have overarching classroom objectives and monitored outliers among students to prevent insufficient averages, they did not establish individual goals for each student in their normal practices. One teacher noted that this approach prompted more consideration for the capabilities and needs of average-performing students *"You know, what really appealed to me is that you're paying more attention to the people in the middle now. We often focus on the stragglers and the high achievers, but that middle*

group sometimes gets overlooked. Now, we're really considering what they can achieve, and I think that's fantastic."

The goal-setting procedure of teachers followed a common structure. The teachers established a baseline using the test scores obtained from this study. Expected improvements were then projected using factors such as previous grades, in-class behaviour, study habits, and pertinent student characteristics, among which intelligence, discipline, work ethic, ability to concentrate, and test anxiety. Some used a balanced representation of variables to establish the goal exemplified by the following excerpt *"Take, for instance, number 12. He scored a 4.5, but I'm aware that he usually achieves good grades and studies diligently. He's inquisitive, asking many questions... a genuinely pleasant student. Therefore, I believe he can attain a 7.5."* Others relied more strongly on a single factor such as previous grades "*Yes, I have definitely looked at their grades (*Referring to students' current average). *If they currently have an 8 or 9, I do expect that they can achieve similar results here"* Notably, the majority of teachers adopted a minimum goal of 5.5 "…*everyone should be able to attain a passing grade"*.

All the participant in this group believed their goals to be realistic and relevant. Teachers perceived the goals as relevant because the goals aligned with a multi-year view of the curriculum. The content that needs to be repeated to attain the goals is part of the final examinations conducted during the last year of high school. *"All of it is exam material, so yes, they have to get to work, and by the end of the year, they should score a couple of points higher."*

One teacher noted an extra layer of relevance, asserting that working on these goals would expose students to the repercussions of last-minute cramming such as incomplete learning, and reduced retention in long-term memory. Another teacher expressed the high experienced relevancy of testing and aiming to improve the long-term comprehension of

previously covered content with the following quote *"I certainly think it is of value, I actually want to test this in my other classes in the same manner"*

The experienced relevancy aligned with a general desire and motivation expressed by the teachers to achieve the set goals. Teachers mentioned being motivated by unexpectedly low test results, the added academic value for students, and the novelty and challenge presented by the context of this study. However, one participant noted that the current academic pressures on students were already high, stating that while the results were interesting, investing considerable effort in pursuing the goals might not be in the best interest of the students due to the risk of overstraining the students.

Goal-first Group. The Goal-First Group expressed a generally positive sentiment toward the goal-first method. Teachers appreciated the challenge it posed to their preconceived notions, one describing the unexpectedly low grades achieved on the test as a sort of wake-up call. Teachers noted that the approach felt more personal, prompting a deeper consideration of the human aspect of students "*I personally found this way of working very interesting, but a bit strange, I needed to switch modes. It makes you think differently, paying more attention to who someone is, how they behave… I think it is more personal*". There were however some drawbacks mentioned, one teacher highlighted the difficulty of establishing fitting goal grades for such a test due to its unfamiliar nature.

The teachers in the Goal-First Group use similar steps to set goals, using multiple factors, including previous grades and general intuition, study habits, and student characteristics, among which natural ability, discipline, and work ethic. The approaches often started with a general impression which was expanded based on one or more factors. The following excerpt illustrates the thought process: "*It* (referring to formulating a goal) *would take about 15 seconds; I would think, where are they at? Additionally, I would consider how they are in class, are they active, do they do their homework?*" Study habits were frequently

referenced, one teacher described it as follows: "*I have really considered who is serious about mathematics, does their homework well, and follows the curriculum."* Other relevant characteristics were also considered, characteristics mentioned by teachers including natural ability, discipline, and attendance. Previous grades were also mentioned by teachers but often with general statements such as "*I have certainly looked at last year's grades".* Although not explicitly discussed in the interviews, the data indicates that most teachers refrained from setting any goals below a passing grade (5.5).

The participants in this group viewed their goals as highly ambitious and indicated the goal differentials were larger than expected. One teacher indicated a preference for slightly lower goals in light of the test results. Despite the challenges, they unanimously believed that with sufficient effort, the set goals were realistic, even if the differentials were substantial. As one teacher put it, *"Yes, 3 or 4 points is a lot, but if they put in the work, I think it should be possible."* All of the teachers expressed motivation to work towards the goals seeing them as relevant and aligning well with the curriculum, often citing the low-test results, indicated by statements such as "*I was startled by it* (the low grades), *I would still like to explore how this can be improved with my class".* The teacher often linked the low grades to bad study habits, such as last-minute cramming for tests and not taking homework seriously. Hence, in addition to academic performance, most of the teachers expressed an additional aim to facilitate the development of better study habits. Teachers earnestly desired students to improve their attitudes towards studying and adopt more consistent study behaviour. The following quotes illustrate this: *"It must really get into their heads that you can't just take a week and study for the test; this just won't get it into long-term memory. They must keep up, keep up, keep up…" and "really show them that it is a waste how much knowledge does not stick…"* (referring to the consequences of intensive last-minute studying*.)*

Comparative Analysis. The Data-First group experienced their approach as similar to normal practices although more comprehensive demanding a more even distribution of attention among students regardless of proficiency, the Goal-First Group found their approach novel and interesting. Stating the goal-first approach potentially led to a more personal assessment of student capacities, focusing on individual characteristics and personal aspects instead of numerical data.

Both groups had distinct starting points: the Data-First Group started their goal-setting process with the study's test data, whereas the Goal-First Group, lacking this data, began with a general impression rooted in intuition. Both groups employed a mix of additional variables to expand on these foundations. The Goal-First Group adjusted to the absence of test data by placing more emphasis on study habits and student characteristics. Notably, a minimum goal of 5.5 was employed by most participants in both groups. Both groups emphasised the importance of improving study habits and integrating them into their goals.

Considering attitudes, the Data-First Group viewed their goals as slightly more realistic, while the Goal-First Group perceived them as more ambitious. Both groups noted the relevance of their goals, emphasising the alignment with the curriculum in the coming years, specifically the final exams. Motivation was a shared sentiment in both groups, with low grades serving as a primary motivator.

Action plans

Data-First Group. Members of the Data-first group used a combination of factors when formulating action plans. First, they strongly relied on test data generated through this study, with many teachers diligently recording the frequency of students who correctly answered specific questions and using this information to identify areas of improvement. *"I don't get it. We recently went over finding the intersection point, and still, almost half got it wrong. Well, we definitely need to spend more time on that."* Previous assessment data was

also used to gauge proficiency levels in the past, teachers often assumed that past proficiency and time needed to relearn a topic were negatively correlated, with high previous proficiency leading to short relearning times.

The second input was personal experience, with teachers often implementing instructional methods and techniques that had proven effective in the past. However, sometimes experience seemed to serve as a justification for including elements in the absence of an immediate substantive reason. As seen in the following quote "*Well, why do I do it like that? There you're asking me something... I think it also stems a bit from experience."*

Lastly, student input was frequently used. Student input has a dynamic function, with this data often becoming available during implementation. Teachers would outline activities, but the content of the activities would often hinge on student input, which generally fell into two broad categories: direct input, such as students identifying challenging questions from the test or topics from classes for discussion, and indirect input, such as observing classroom dynamics and atmosphere during specific topics.

Based on these inputs action plans were formulated in which the Data-First Group focussed on aspects of instruction, assessment, and curriculum. A primary challenge acknowledged by the group was time constraints, emphasizing that the current curriculum provided minimal flexibility. In terms of instruction, there was a diversity of strategies among teachers. One teacher planned to use extra time at the end of the year for a general review of materials. *"In the schedule, I have 2 or 3 classes planned in case I get sick or something, I thought, what if I use them to quickly go over the material again. Of course, If I don't actually get sick (laughs)."*

A second teacher wanted to keep the previous topics salient through homework and homework discussion "*I was thinking of tackling it through homework. Just adding an*

assignment from a previous chapter every week. That way, during the homework discussion, we can also go over that topic if needed."

A third teacher used two methods, starting with enforcing homework more strictly to stimulate better studying behaviour "*I do want to start checking homework more strictly again, although I always find that a bit patronizing. But now you can see again how important it is for them to keep up*." Additionally, this teacher planned to use the remaining lessons as an opportunity to rehearse for a second assessment "*Towards the end of the year, I usually have one or two extra lessons left. If that's the case again this year, I'm thinking of using one for questions and the other for the test."* The fourth teacher emphasized personal responsibility, suggesting that students should review the materials and prepare for a new assessment independently, without additional support or instruction. *"I don't see a possibility to dedicate extra time to it; it will have to be done in their own time."*

The teachers tried to gain additional time and resources were possible, one teacher aimed to explore whether some of the time from cancelled lessons throughout the year could be used to focus on mathematics homework.

Regarding assessment, three teachers in this group preferred a single, comprehensive assessment at the year's end. In contrast, one teacher was inclined to integrate earlier topics into ongoing tests. While one teacher planned to inform students about these additional topics, another preferred not to specify which topics would be repeated, ensuring students would have to go over all the materials "*I think it is a good idea to have a kind of test at the end of the year, like an exam, which can contain any of the topics covered that year."*

There is variation in how teachers differentiate among students, however, not all teachers articulated a differentiation strategy. The teachers seemed to have preexisting practices for differentiating, which were then applied to the action plan. The discussed strategies encompass permitting proficient students to progress independently setting their

own pace, organizing groups of similar proficiency levels for collaborative assignments, and forming groups inclusive of diverse ability levels "*A few of them grasp it very easily. Sometimes, I ask one of them to collaborate with someone who finds it a bit challenging... I think they can both learn from that; explaining can also be very instructive"*

Goal-First Group. The Goal-First group employed test data extensively, with most teachers maintaining comprehensive records detailing individual student performance on specific questions. This meticulous approach was evident not only in interviews but was further substantiated by some participants who shared anonymized datasheets reflecting this practice. Student input held significant importance, with several teachers using comprehensive post-test discussions to gain insight into the learning needs of students *"...after the test, I divide the students into 2 or 3 groups, and then each group has to discuss which questions were the most difficult. We then go over those questions together on the board."*

Prior experience was also a common factor, often employed to justify one's methods and strategies. Most relied on previous teaching experience, but one teacher recounted a memorable practice from their student days. *"Oh, I used to have a teacher who always asked about the colours of the rainbow in the tests, and that has always stuck with me, you know? So, I thought, why not try something similar? Bringing back old topics in the tests so that students keep reviewing and remembering what they've learned."*

The participants used a range of instruction and assessment strategies in combination with differentiation strategies. The teachers used strategies for instruction such as revisiting materials, focusing on improving learning habits and giving additional homework. To implement this some planned to modify existing lessons while others allocated entire lessons, one teacher was very proactive in this regard as illustrated by the following quote "*Next week I've already freed up a class, I want to go over the test together, especially those questions*

where almost the entire class made mistakes. Numbers 3 and 9, if I recall correctly." Additionally, most teachers wanted to emphasize improving study habits. One teacher expressed it as follows: *"I've always struggled to change my student's study habits, but I'm ready to try. Maybe this test is the proof I need to show how wasteful this way of studying is".* To achieve this another teacher mentioned the following plan: *"I was thinking of maybe doing a couple video meetings in the evening the coming months where students can work together on their homework with me."*

To practice the materials the teachers had multiple strategies, on one wanted more integral questions covering a broader range of topics simultaneously, another teacher proposed a weekly review of old test problems. *"First, I want to discuss the test results with students... But I thought it might be fun to do a weekly problem, maybe from old tests, on the whiteboard together."* Additional homework was only considered by one, who stated it was an option but only if he could not manage to fit the practice inside the lessons.

Regarding assessment, one teacher proposed a "*20/80 structure*" suggesting that 20 percent of the questions would address previous test materials, while 80 percent would focus on current materials. Additionally, one teacher aimed to include a reflective component in tests, encouraging students to ponder over the problems and their study methods. "*I haven't got the other teachers on board yet, but I'd like them* (referring to the students) *to reflect at the end on how they've learned, whether they've done the homework, and if there's anything they would do differently... For a small bonus."*

Most teachers in this group were keen on implementing some form of student differentiation. The teachers used the strategies from their normal teaching practice in the action plans. These strategies included enabling proficient students to progress ahead of the class, providing students with the option to work independently or collaboratively with the

teacher, and forming heterogeneous groups comprising students of varying ability levels at the outset of complex topics.

Comparative Analysis. The data-first and goal-first approaches relied on similar information when formulating the action plans. In both groups, teachers often kept extensive records of how the students performed on the provided test. Using this information to indicate areas of improvement. Student input was used to clarify the learning needs of students and adjust the curriculum accordingly. Furthermore, the teacher often relied on their experience, implanting strategies that had been successful in the past. The action plans of both groups can be described based on instruction and assessment. Both groups used a range of instructional strategies often implementing a range of strategies focussing on homework and lesson content. The data-first group saw time constraints as a big object, more often focussing on homework and only using the lesson time at the end of the year if possible. One teacher did not make any active changes to instruction leaving the choice and responsibility to students. However, others did aim to allocate lesson time in the near future, one teacher aimed to allocate lesson time for homework discussion throughout the year, while another sought to utilize the time from cancelled lessons for math homework.

The goal-first group seemed to be more willing to use lesson time throughout the year and focussed less on homework, aiming to work on assignments in the classroom setting. This group also allocated time to focus on improving study habits, with one teacher even aiming to provide separate sessions for this.

Assessment strategies were aligned with instructional strategies, with the data-first group focusing primarily on end-of-year assessments. In contrast, the goal-first group aimed to implement assessments more consistently throughout the year, often opting to modify existing assessments. However, both groups utilized both approaches to some extent. Both

groups employed a range of differentiation techniques, in both groups this was a clear extension of the teacher's normal practices.

Teacher Efficacy

This section pertains to RQ 3 "*How does teacher efficacy influence goal differentials and respective action plans?"* To this end, the three dimensions of efficacy are discussed concerning action plan contents followed by a short discussion of general efficacy and goal differentials.

Teacher Efficacy in Instructional Strategies. Teacher efficacy in instructional strategies seems similarly distributed between the two groups (data-first and goal-first). In both groups, half of the teachers gave answers which indicated high efficacy, while the other half indicated medium efficacy. Note that these are not comprehensive measures but interview answers.

Teachers who demonstrated high efficacy in instructional strategies showed a tendency to differentiate more and employ more innovative and engaging methods in their action plans compared to medium efficacy teachers. This is illustrated most clearly by quotes from teacher G from the goal first group who expressed high efficacy with the statement: "*To me, it is crucial to tailor my teaching to student's needs, especially in mathematics where students tend to get easily discouraged by failure. Aligning with their current level makes a huge difference, offering them the experience of success."* This teacher demonstrated strong differentiated education: *"I always try to align as much as possible with individual learning needs. Some children work independently, even determining what they do in a lesson themselves because they understand it on their own. However, with others, I sit with them in almost every lesson to work on assignments together.*" Finally, this teacher showed an innovative yet engaging instructional method: "*But what I thought might be a good idea every week is to take an old assignment, perhaps from a previous test, and work it out together on the board."*

Teacher Efficacy in Classroom Management. In the realm of classroom management, there was minimal differentiation among the group. All teachers, except one, expressed a strong sense of control within their classrooms; hence, no distinct patterns could be observed in the action plans based on efficacy levels in this category.

Teacher Efficacy in Student Engagement. Two teachers in the data-first group and one in the goal-first group expressed a strong efficacy belief for student engagement. These teachers indicate the importance of using pedagogical strategies that foster engagement and curiosity. This aligns with Teacher E's perspective, as captured in the following quote: "*I notice that simply showing how something works is often much better than explaining it. When I used to explain conversions in primary school, I always brought my box of 10 by 10. You know that a litre is a cubic decimetre, right? Then I would always ask, who wants to bet that a litre fits in there. Just little things that capture the imagination."*

Furthermore, teachers who felt confident in their ability to engage students also tended to value student input in their lessons. This is exemplified by the practices of Teacher B and G: "*I divide the students into 2 or 3 groups after the test, and then each group has to discuss which questions were the most difficult. We then go through those together on the board."* and "*I prefer not to plan these things too detailed; I believe you should gauge the learning needs in the class and factor that into how much time you spend on topics."*

Teacher Efficacy and Goal Differentials. The interviews showed no clear relationship between teacher efficacy and goal differential. Participants E and G, displaying high efficacy across all three dimensions, also held the second and third-highest goal differential scores. However, participant F, with the highest goal differential, indicated high efficacy in only one category. The remaining participants did not exhibit a clear pattern between goal differentials and efficacy. It is essential to acknowledge the inherent subjectivity

in evaluating efficacy and the lack of a discernible pattern among participants in terms of goal differentials and efficacy.

Discussion

The primary objective of this study was to enhance the understanding of the relationship between Data-Informed Decision Making (DIDM) approaches and academic outcomes in the form of grades. Specifically, this study aimed to determine whether the goal-first approach or the data-first approach would result in higher goals and higher-quality action plans. Additionally, the study explored the moderating influence of teacher efficacy on the relationship between DIDM approaches and educational outcomes.

Main Findings

DIDM Approaches and Goal-differentials.

The analysis of DIDM approaches and goal differentials leaves us with inconclusive evidence. On average, the goal-first group exhibited a goal differential 0.84 units larger than the data-first group. Interestingly, initial analyses at the teacher level suggested this disparity was not statistically significant, whereas subsequent analyses at the student level indicated the goal-first approach resulted in significantly higher goal differentials. To reconcile these discrepancies, a linear mixed model was applied, encompassing both group and student levels. However, this extensive examination failed to confirm the relationship after accounting for teacher-level variability, resulting in a *p-value* of 0.108. The limited number of teachers (N= 9) may have contributed to insufficient statistical power, thus hindering the identification of a significant pattern.

DIDM Approaches, Goal Setting and Action Plans

The central theme in the goal-first approach, besides attaining the goal level, seems to be improving long-term comprehension. In the interviews, the teacher expressed enthusiasm about how the goal-first approach prompted them to thoroughly consider the personal characteristics and behaviours of students. During goal setting, in the absence of the test

grades, the teachers relied strongly on student study habits, linking last-minute cramming and poor homework practices with lower comprehension and expectations.

The importance of study habits was emphasized further when the teachers were confronted with the low grades, revealing most of the students were not effectively retaining learned materials. This connection between study habits and long-term comprehension is supported by the literature which consistently demonstrates that cramming, characterized by intense, last-minute study sessions, is associated with poor long-term retention of information and skills (McIntyre & Munson, 2008). Study habits became a central part of the action plans, with some teachers even scheduling separate sessions for this sole purpose. The teachers planned to emphasize study habits using various tools such as classroom discussion about the test and its results and modifying future assessments to include a section about previously covered materials to ensure these materials remain salient. Kang (2016) highlights the effectiveness of such spaced repetition, distributed over an extended period, in enhancing long-term retention compared to massed, all at once, practice. Kang's research shows the effects of spaced repetition have been consistently confirmed by hundreds of experiments.

In the data-first approach, concerns regarding long-term comprehension were also acknowledged. Teachers observed low grades and recognized the need for practices that encourage better study habits. However, the primary focus of this group was achieving goals. Time was seen as a significant obstacle, prompting a focus on efficiency. Consequently, this group often wanted a single assessment and emphasized tools such as homework and using buffer days to manage time constraints effectively.

However, the effectiveness of homework in this context is not certain. While Kitsantas et al. (2011) suggest that homework can improve understanding of mathematics, they stress the significance of having accompanying instruction. If students are not given structured examples and explanations beforehand, and if they do not discuss the materials afterwards, it

is unlikely to stimulate comprehension in a meaningful way. Furthermore, a study conducted in Dutch high schools casts doubt on the effectiveness of increasing homework in high school, as it found no significant correlation between homework frequency and grades while the volume of homework only explained a marginal amount of variance (de Jong et al., 2000).

DIDM Approaches and Attitudes

The goal-first group tended to perceive their goals as highly ambitious, with many stating the goal differentials were larger than expected. Conversely, among the teachers in the data-first group, only a portion characterized their goals as ambitious, and when these characterizations were made this was often done with general statements indicating a need to get to work. This discrepancy in the content and frequency of statements regarding ambitiousness suggests that the goal-first group perceived their goals as more ambitious. The relevance of this difference is underscored by Locke and Latham (2019), who demonstrated a positive correlation between the ambitiousness of goals and the level of commitment individuals exhibit toward achieving them, provided that the goals remain attainable. While acknowledging the significance of this difference, it is important to highlight that the remaining attitudes toward the goals exhibited notable similarities between the groups.

With few exceptions, the majority of teachers in both groups perceived the goals as relevant and realistic, while also expressing a strong sense of motivation to achieve them. The uniformly positive attitudes of participants should be cautiously considered due to the potential influence of self-selection bias. Self-selection bias can occur because participants who choose to engage in research may not represent the broader population in a field accurately in every respect. For example, high-functioning professionals are often overrepresented in research samples using self-selection (Elston, 2021).

The Role of Efficacy

The distribution of efficacy levels appears relatively even within the goal-first and data-first groups, suggesting it did not have an important impact on the analysis of goal differentials between the groups. The connection between teacher efficacy and goal differentials remains unclear, although there is a slight indication that high efficacy across all three dimensions might be associated with high goal differentials, other studies support this possibility (Allinder, 1994; Ross & Bruce, 2007).

Furthermore, high efficacy in student engagement and instructional strategies appears to impact the instructional strategies teachers apply, with the higher efficacy teachers emphasising the importance of instructional strategies that foster engagement and curiosity. This observation aligns with existing literature, which suggests that teachers with high efficacy demonstrate a willingness to explore innovative and challenging teaching ideas and techniques (Ross, 1998; Allinder, 1994). Consequently, there are indications that heightened teacher efficacy positively influences the quality of action plans, goal setting, and ultimately education. Hence the tentative suggestion that, in the context of training teachers using DIDM methods, some consideration should be given to enhancing efficacy beliefs.

Recommendations

While this study cannot definitively state which DIDM approach is most effective, it does showcase the potential positive effects of implementing DIDM approaches in educational settings. Both groups of teachers noted that using a DIDM approach had prompted them to apply a more comprehensive approach to goal setting, distributing attention more equally among students and considering different aspects, such as study habits, more seriously. Ultimately, this resulted in two groups of motivated teachers, who were capable and driven to enact change. Previous research also shows that DIDM can have improved educational results (Van Geel et al., 2016). Hence, it is strongly recommended educational institutions implement some form of DIDM when possible.

While both approaches exhibit merit, the data from this study appears more favourable towards endorsing the goal-first approach. However, it is important to acknowledge the nuanced nature of this assertion, as definitive conclusions cannot be drawn, and various contextual factors may influence the suitability of each approach in specific circumstances.

Moreover, it is crucial to ensure that teachers are willing and capable of implementing these approaches effectively. Research by Keuning et al. (2017) highlighted the importance of teachers having positive attitudes toward data-use methods and possessing adequate data-use skills. Teachers with positive attitudes toward data and higher literacy skills were more likely to achieve favourable outcomes with DIDM interventions. Therefore, when implementing DIDM approaches, it is essential to provide support and training to ensure that teachers have the necessary skills and attitudes to effectively use data in decision-making processes. Promoting teacher efficacy appears to align with these objectives, as higher efficacy and positive attitudes towards data and data literacy are correlated (Dunn et al., 2013). Additionally, this study, along with other research, suggests that teachers with higher efficacy levels tend to excel employ high-quality goal attainment strategies such as implementing personalized and creative methods for instruction.

While not the primary focus, this study shed light on the critical issue of poor long-term retention of mathematical concepts among high school students. Teachers' observations underscored its significance, emphasizing the need for comprehensive assessments, particularly focusing on long-term comprehension and skill retention. Incorporating or modifying assessments to measure this can provide educators with a better understanding of retention levels and inform adjustments to instruction.

Limitations

While this research paper provides interesting results that prompt further investigation of DIDM, and educational outcomes certain limitations should be considered. Due to time constraints, action plans and goal differentials were employed as indicators to gauge future

educational improvement. While these indicators are valid, they can be subject to error and potential inaccuracies. For instance, the implementation of action plans frequently introduces unforeseen challenges stemming from disparities between planned and real-world realities (Schildkamp, 2019). Therefore, although the gathered insights are pertinent, it is beneficial to seek additional direct measures such as grade improvement over time on comparable tests.

This study faces constraints due to the limited sample size. The recruitment of participants was challenging because teachers needed to allocate lesson time for test administration and a general sense of research exhaustion among schools. Ultimately, these factors resulted in the gathering of a relatively small group of participants, impacting the overall generalizability and reliability of the study.

Lastly, the test and testing procedure were different from normal mathematic test practices, this potentially had an impact on the observed goal differentials. The test used in this study assessed the mathematical knowledge of students on a broad range of previously covered materials and was administered without prior preparation. In contrast, typical mathematics assessments concentrate on a narrower scope of topics which have typically been discussed in class recently and allow students to prepare beforehand. Because teachers have little to no experience with this type of test situation their intuitions about student performance are likely less accurate (Betsch, 2008; Klein, 2017). During the interview participants confirmed this, stating the test results were lower than expected. Consequently, it is plausible that the goal-first group's differentials turned out higher than normal, as the starting point was lower than expected, resulting in a greater distance to the goal level. In contrast, the data-first group, being informed about grades before setting goals, likely remained unaffected by this influence. However, it is noteworthy that the test received a highly positive reception from the teachers due to the unique insights it offered into the long-term retention of materials.

Future Research

The findings of this study are inconclusive but suggest a possible positive relationship between the goal-first approach and goal differentials highlighting the need for additional research with a larger sample size to validate or reject this relationship. Furthermore, future research should conscientiously assess the test type and method employed; it is advisable to consider implementing a more standard test, preferably in a scenario closely resembling the typical testing conditions in a classroom setting. This approach aims to enhance generalization and mitigate the potential of the test influencing teacher intuitions.

Future research on the impact of DIDM approaches would benefit from longer timeframes so direct measures of educational outcomes can be taken. A more prolonged study duration will provide a clearer demonstration of relationships between DIDM approaches and factors such as goal setting, grades and action plans and help clarify whether elements like experience with different methods or data influence these relationships.

Finally, the linear mixed model employed in this study suggests a comprehensive understanding of the impact of data-use interventions on academic performance including an understanding of teacher variables, namely individual characteristics of teachers. Conducting a study with a within-subjects design, in which all teachers set goals using both methods could shed light on the impact of teacher variables on the effectiveness of both approaches.

Conclusion

While this study provides valuable insights into the impacts of the goal-first and data-first DIDM approaches within educational settings, no definitive conclusion can be drawn regarding the superiority of either approach in the context of action plans and goal-setting. Ultimately, the significant positive relationship between goal differentials and the goal-first group could not be confirmed when accounting for teacher-level variations, possibly due to the limited number of participating teachers. While the current data does not definitively

establish the relationship, suggestive trends allow for optimism regarding the confirmation of this relationship in subsequent research.

Furthermore, while the goal-first approach emphasized long-term comprehension through adjustments in study habits, assessments, extra lessons, and other available resources, the data-first approach prioritized goal attainment with minimal effort, relying more on personal responsibility and resource utilization such as homework and remaining lesson time, if feasible.

Additionally, while both groups displayed largely positive attitudes towards their goals, the goals of the goal-first group stood out notably. These goals are closely aligned with the literature on optimal goal-setting, being ambitious, attainable, and relevant (Locke & Latham, 2019). Such goals inspire teachers and foster an environment conducive to student development.

These findings add to the expanding body of research on DIDM, particularly about comparing different approaches—a topic that has been scarcely explored in the literature. This study offers an initial insight into the comparative effects of the goal-first and data-first methods and lays the groundwork for future research.

Bibliography

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education*, 17(2), 86-95.

Babbie, E. R. (2021). *The Practice of Social Research* (15th ed.). Cengage.

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2020). Big data in education: A state of the art, limitations, and future research directions. *International Journal of Educational Technology in Higher Education*, 17, 44. https://doi.org/10.1186/s41239-020-00223-0

Baron-Thiene, A., & Alfermann, D. (2015). Personal characteristics as predictors for dual career dropout versus continuation—A prospective study of adolescent athletes from German elite sport schools. *Psychology of Sport and Exercise, 21,* 42–49. https://doi.org/10.1016/j.psychsport.2015.04.006

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122-147. https://doi.org/10.1037/0003-066X.37.2.122

Bandura, A. (1988). Self-Regulation of Motivation and Action Through Goal Systems. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), Cognitive Perspectives on Emotion and Motivation (NATO ASI Series, Vol. 44). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-2792-6_2

Betsch, T. (2008). The nature of intuition and its neglect in research on judgment and decision making. In H. Plessner, C. Betsch, & T. Betsch (Eds.), *Intuition in Judgment and Decision Making* (pp. 3–22). New York, NY: Lawrence Erlbaum Associates

Burla, L., Knierim, B., Barth, J., Liewald, K., Duetz, M., & Abel, T. (2008). From text to codings: intercoder reliability assessment in qualitative content analysis. Nursing research, 57(2), 113-117

Button, K., Ioannidis, J., Mokrysz, C., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376. https://doi.org/10.1038/nrn3475

Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability, 21*(1), 47–65. https://doi.org/10.1007/s11092-008-9063-x

Carvalho, M., Cabral, I., Verdasca, J., & Alves, J. M. (2022). Strategic Action Plans for School Improvement: An Exploratory Study About Quality Indicators for School Improvement Plan Evaluation. Journal of Social Studies Education Research, 13(1), 143-163.

De Jong, R., Westerhof, K. J., & Creemers, B. P. M. (2000). Homework and student math achievement in junior high schools. *Educational Research and Evaluation*, 6(2), 130-157. https://doi.org/10.1076/1380-3611(200006)6:2;1-E;F130

De Vries, J. A., Dimosthenous, A., Schildkamp, K., & Visscher, A. J. (2022). The impact on student achievement of an assessment for learning teacher professional development program. *Studies in Educational Evaluation*, 74, 101184. https://doi.org/10.1016/j.stueduc.2022.101184

Dunn, K. E., Airola, D. T., Lo, W.-J., & Garrison, M. (2013). Becoming Data Driven: The Influence of Teachers' Sense of Efficacy on Concerns Related to Data-Driven Decision Making. *The Journal of Experimental Education, 81*(2), 222-241. https://doi.org/10.1080/00220973.2012.699899

Earl, L. & Fullan, M. (2003). Using Data in Leadership for Learning. *Cambridge Journal of Education*. 33. 383-394. 10.1080/0305764032000122023.

Elston, D. M. (2021). Participation bias, self-selection bias, and response bias. *Journal of the American Academy of Dermatology.* Advance online publication. https://doi.org/10.1016/j.jaad.2021.06.025.

Grabarek, J., & Kallemeyn, L. M. (2020). Does Teacher Data Use Lead to Improved Student Achievement? A Review of the Empirical Evidence. *Teachers College Record, 122*, 42.

Hollenstein, L., Rubie-Davies, C. M., & Brühwiler, C. (2023). Teacher expectations and their relations with primary school students' achievement, self-concept, and anxiety in mathematics. *Social Psychology of Education, 27*(4), 567–586. https://doi.org/10.1007/s11218-023-09856-1

Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "Data-Driven" Mantra: Different Conceptions of Data-Driven Decision Making. *Teachers College Record*, 109(13), 105–131. https://doi.org/10.1177/016146810710901310

Jimerson, J. B. (2014). Thinking about data: Exploring the development of mental models for data use among teachers and school leaders. *Studies in Educational Evaluation*, 42, 5-14.

Kahneman, D., & Frederick, S. (2005). A Model of Heuristic Judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge University Press.

Kang, S. H. K. (2016). Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. *Policy Insights from the Behavioral and Brain Sciences, 3(1), 12-19. https://doi.org/10.1177/2372732215624708*

Keuning, T., Van Geel, M., & Visscher, A. (2017). Why a Data–Based Decision–Making Intervention Works in Some Schools and Not in Others. Learning Disabilities Research & Practice, 32(1), 32-45. https://doi.org/10.1111/ldrp.12124

Kitsantas, A., Cheema, J., & Ware, H. W. (2011). Mathematics achievement: The role of homework and self-efficacy beliefs. Journal of Advanced Academics, 22, 310–339.

Klein, G. (2017). *Sources of power: How people make decisions*. MIT Press.

Krein, U., & Rohs, M. (2021). Data in Schools: (Changing) Practices and Blind Spots at a Glance. Frontiers in Education, 6. https://doi.org/10.3389/feduc.2021.672666.

Lai, M., & Schildkamp, K. (2013). Data-based decision making: An overview. In K. Schildkamp, M. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities*. Studies in educational leadership, vol. 17 (pp. 9–21). Springer. https://doi.org/10.1007/978-94-007-4816-3

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance. *Prentice -Hall, Inc*.

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science*, 15(5), 265-268.

Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science*, 5(2), 93.

Mandinach, E. B. (2012). A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. *Educational Psychologist.* 47. 71-85. 10.1080/00461520.2012.667064.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30-37.

Marsh, J. A. (2012). Interventions Promoting Educators' Use of Data: Research Insights and Gaps. In *Teachers College Record* (Vol. 114).

McIntyre, S. H., & Munson, J. M. (2008). Exploring cramming. *Journal of Marketing Education*, *30*(3), 226–243. https://doi.org/10.1177/0273475308321819

Opper, I. M. (2019). Teachers Matter: Understanding Teachers' Impact on Student Achievement. RAND Corporation. https://www.rand.org/pubs/research_reports/RR4312.html

Poortman, C. L., & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and teacher education*, *60*, 425-433. https://doi.org/10.1016/j.tate.2016.06.010

Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. In J. Brophy (Ed.), *Research on Teaching* (Vol. 7, pp. 51-71). Simon & Schuster Macmillan.

Ross, J., & Bruce, C. (2007). Professional development effects on teacher efficacy: Results of a randomized field trial. *The Journal of Educational Research*, 101(1), 50-60. Retrieved from ERIC database. (ERIC Document Reproduction Service No. EJ 776 276)

Rubie-Davies, C., Hattie, J. and Hamilton, R. (2006). Expecting the best for students: teacher expectations and academic outcomes. *British Journal of Educational Psychology, 76(3),* 429–444.

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. Instructional Science, 18, 119-144.http://dx.doi.org/10.1007/BF00117714

Sarker I. H. (2021). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN computer science*, *2*(5), 377. https://doi.org/10.1007/s42979-021-00765-8

Schildkamp, K. (2019). The goals-first approach to data use in schools. *Journal of Educational Change*, 20(1), 5-28.

Schildkamp, K. & Datnow, A. (2022) When Data Teams Struggle: Learning from Less Successful Data Use Efforts. *Leadership and Policy in Schools*, 21(1), 147-166, DOI: 10.1080/15700763.2020.1734630

Shabani, K., Khatib, M., Tabataba'i Uinversity, A., & Ebadi, S. (2010). *Vygotsky's Zone of Proximal Development: Instructional Implications and Teachers' Professional Development* (Vol. 3, Issue 4). www.ccsenet.org/elt

Shorten, A., & Smith, J. (2017). Mixed methods research: Expanding the evidence base. *Evidence-Based Nursing*, 20, 74-75.

Soland, J. (2013). Predicting High School Graduation and College Enrollment: Comparing Early Warning Indicator Data and Teacher Intuition. *Journal of Education for Students Placed at Risk (JESPAR)*, 18(3-4), 233-262. https://doi.org/10.1080/10824669.2013.833047

Timperley, H., & Phillips, G. (2003). Changing and sustaining teachers' expectations through professional development in literacy. *Teaching and Teacher Education*, 19(6), 627-641.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202-248.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17, 783-805.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. In *New Series* (Vol. 185, Issue 4157).

Van den Berg, S. M. (2021). *Analysing Data using Linear Models*. (5 ed.) University of Twente

Van der Scheer, E. A., & Visscher, A. J. (2016). Effects of an intensive data-based decision making intervention on teacher efficacy. *Teaching and Teacher Education,* 60, 34-43. https://doi.org/10.1016/j.tate.2016.07.025.

Van Geel, M. J. M., Keuning, T., Visscher, A. J., & Fox, G. J. A. (2016). Assessing the effects of a school-wide data-based decision making intervention on student achievement growth in primary schools. *American Educational Research Journal*, 53(2), 360-394. https://doi.org/10.3102/0002831216637346

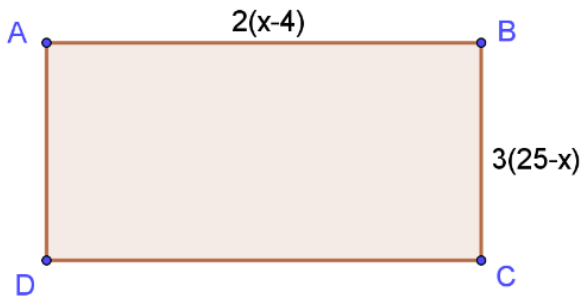
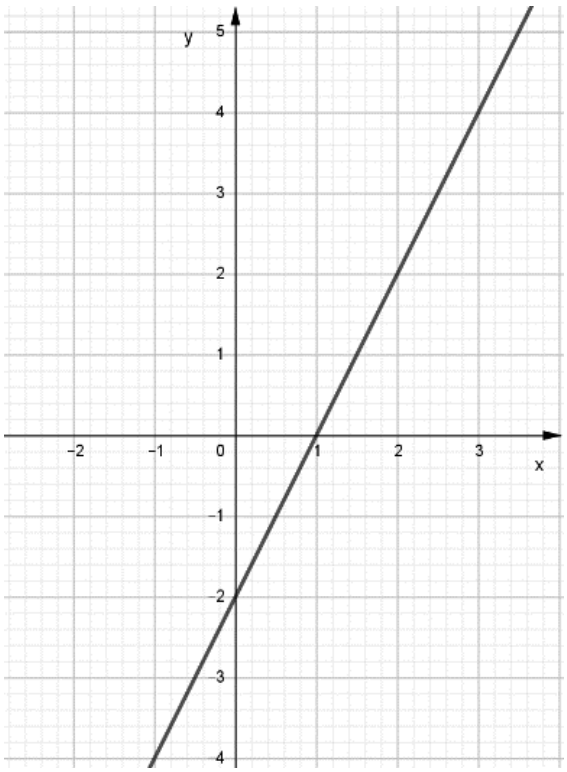
Vanlommel, K., van Gasse, R., Vanhoof, J., & van Petegem, P. (2017). Teachers' decision-making: Data based, or intuition driven? *International Journal of Educational Research, 83,* 75–83. https://doi.org/10.1016/j.ijer.2017.02.013

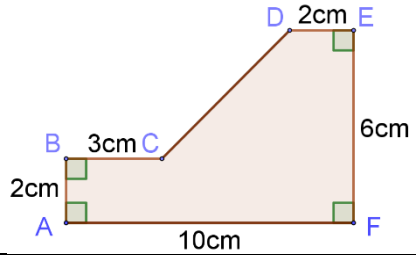
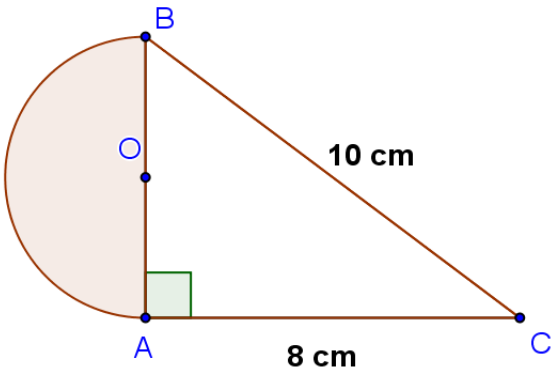
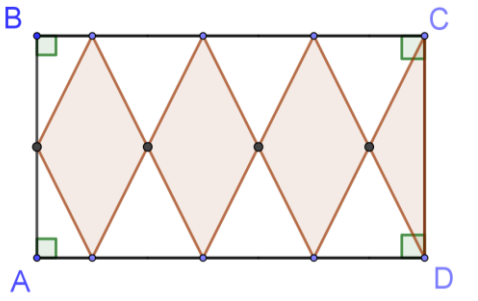
Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 Years. *Educational Research and Evaluation, 24*(3-5), 124–179. https://doi.org/10.1080/13803611.2018.1548798

Appendix I: Answer key test 3 (Dutch)

Answer key test 3 HAVO/VWO in Dutch Modified from de Vries et al. (2022).

Vraag	***Antwoord/ beoordeling***	***Score***
1. *Los onderstaande vergelijkingen op.* a) $6\alpha - (3\alpha - 1) = -2$ b) $x + \frac{2x}{3} = 2$	***1 punt per correct antwoord*** a: $a = -1$ b: $x = \frac{6}{5}$, of $x = 1\frac{1}{5}$, of $x = 1{,}2$ (of een andere gelijkwaardige waarde)	
2. *Werk de haakjes weg en vereenvoudig het antwoord.* $(x-3)(3x^2+x)$	***2 punten voor een correct antwoord*** A = $3x^3 - 8x^2 - 3x$ ***1 punt voor correct maar niet-vereenvoudig Antwoord.*** Bijvoorbeeld: 3x3+x2−9x2−3X	
3. *Voer de berekening uit.* **(a)** $2 \cdot 3^2 - 4\left(3 - 1^{10}\right) + \left(1-2\right)^3 =$ **(b)** $\sqrt{5 \cdot \sqrt{25}} + 2\sqrt{49} - \sqrt{0{,}16} =$	***1 punt per correct antwoord*** a: 9 b: 18,6 , of $18\frac{6}{10}$,	
4. *Het volume van piramide ABCDE kan berekend worden door de formule* $V = \frac{1}{3}a^2h$ ***. Het vierkante grondvlak ABCD van deze piramide heeft een lengte a en de hoogte van de piramide wordt uitgedrukt in h. Het volume van de gehele piramide is*** $V = 75cm^3$ ***en de hoogte is*** $h = 9cm$ ***.*** ***Wat is dan de lengte*** a ***van het grondvlak?***	***1 punt*** $a = 5\text{cm}$ of 5	

5. ***De lengte en breedte van rechthoek ABCD is*** $L = 2(x-4)$ ***en*** $W = 3(25-x)$***. Druk de omtrek van het rechthoek ABCD uit in x.***  	***1 punt*** De omtrek is $-2x + 134$	
6. ***Een groep van 42 mensen (kinderen en volwassenen) maken een busreis. Ieder kind betaalt €5,- voor een ticket, iedere volwassene betaalt €12,-. In totaal werd er €420,- besteed aan de busreis.*** **Hoeveel kinderen gingen er mee op deze busreis?**	***1 punt*** 12 Kinderen	
7. ***De lineaire lijn hiernaast geeft de relatie weer tussen de variabelen x en y.*** **Wat is de richtingscoëfficiënt van de lijn?**  	***1 Punt*** Rc = 2	

8. Wat is de oppervlakte van dit figuur?  	***1 punt*** $38\ cm^2$ *of* 38	
9. Onderstaande figuur bestaat uit de rechthoekige driehoek ABC en de halve cirkel AB met diameter AB. AC = 8 cm en BC = 10 cm. **(a) Bereken de lengte van de zijde AB van de driehoek ABC.** **(b)**  **Bereken de oppervlakte van de halve cirkel** (laat **π** in je antwoord staan).	***1 punt per correct antwoord*** a: AB = 6 of 6 b: $2\frac{9\pi}{2}cm^2$, or $\frac{9\pi}{2}$ Reken tevens goed indien de student de correct formule gebruikt maar op basis van een verkeerd antwoord bij opdracht A verder rekent.	
10. Het rechthoekige tapijt ABCD heeft een patroon van identieke ruiten, zoals je kunt zien in de weergave hieronder. De lengte van het tapijt is 2.80 meter en de breedte is 1.50 meter. Wat is de totale oppervlakte van de ruiten op het tapijt?  	***1 punt*** $2{,}1\ m^2$ of $2{,}1$	

Totaalscore =

Cijfer (Score x 0,643 + 1 =

Appendix II: Interview Schema (Dutch)

Introductie:

- Voorstellen
- Doel: het verkrijgen van inzichten over ervaringen tijdens een onderzoek.
- Privacywaarborging: gegevens worden geanonimiseerd.
- Verwijdering van de transcriptie na de opname.
- Gebruik van gegevens voor analytische doeleinden.
- Geschatte duur van het interview: 40 minuten.
- Delen van resultaten via e-mail.
- Focus op ervaringen en inzichten met betrekking tot het doelstellingsproces van docenten.
- Start van de opname van het interview.

Hieronder volgt het interviewschema, schuingedrukte vragen zijn suggesties voor eventuele doorvragen om aanvullende informatie te vergaren.

Doelen

- Ik heb u gevraagd om voor iedere leerling een doel in de vorm van een cijfer op te stellen. Op basis waarvan heeft u deze doelen opgesteld?
- Welke rol speelden de toetsresultaten bij het opstellen van deze doelen? (Alleen data-first)
    - *Heeft het voordelen om eerst naar de resultaten te kijken en dan doelen op te stellen? Zo ja, welke?*
    - *Heeft deze manier van werken ook mogelijke nadelen? Welke?*
- Ik heb u gevraagd om een doel op te stellen voordat u de resultaten van de leerlingen op de toets had gezien. Hoe vond u dit? (Goal-first)

        - *Had u liever de resultaten op de toets gezien voordat u de doelen ging opstellen? Zo ja, waarom? Zo nee, waarom niet?*
        - *Heeft het voordelen om eerst de doelen op te stellen en dan pas naar de resultaten te kijken? Zo ja, welke?*
        - *Heeft deze manier van werken ook nadelen? Welke?*
- Kunt u mij meenemen in uw gedachtegang bij het opstellen van de doelen?
    - *Welke factoren heeft u meegenomen in uw overweging?*
    - *Kunt u mij meenemen in de gedachtegang bij het doel van een specifieke leerling (zonder zijn of haar naam te noemen)?*
- Hoe realistisch zou u de door u gestelde doelen noemen?
    - *Als u de data had gehad voordat u de doelen opstelde, denkt u dat u dan andere doelen hebben gesteld? (Alleen goal-first)*
- Hoe ambitieus zou u de door u gestelde doelen noemen?
- In hoeverre sluiten de doelen aan op de geplande leerlijn?
- In welke mate voelt u zich gemotiveerd om met dit doel aan de slag te gaan?

Actieplan

- Hoe zou een actieplan om deze doelen te realiseren eruitzien?
- Waar is dit actieplan op gebaseerd? Doorvragen op:
    - Eigen ervaring?
    - Data
    - Wetenschappelijke kennis en literatuur
- Welke concrete aanpassingen zou u maken in het curriculum?
    - *Welke aanpassingen in instructie?*
    - *Welke aanpassingen in assessment?*
    - *Welke aanpassingen in opdrachten?*

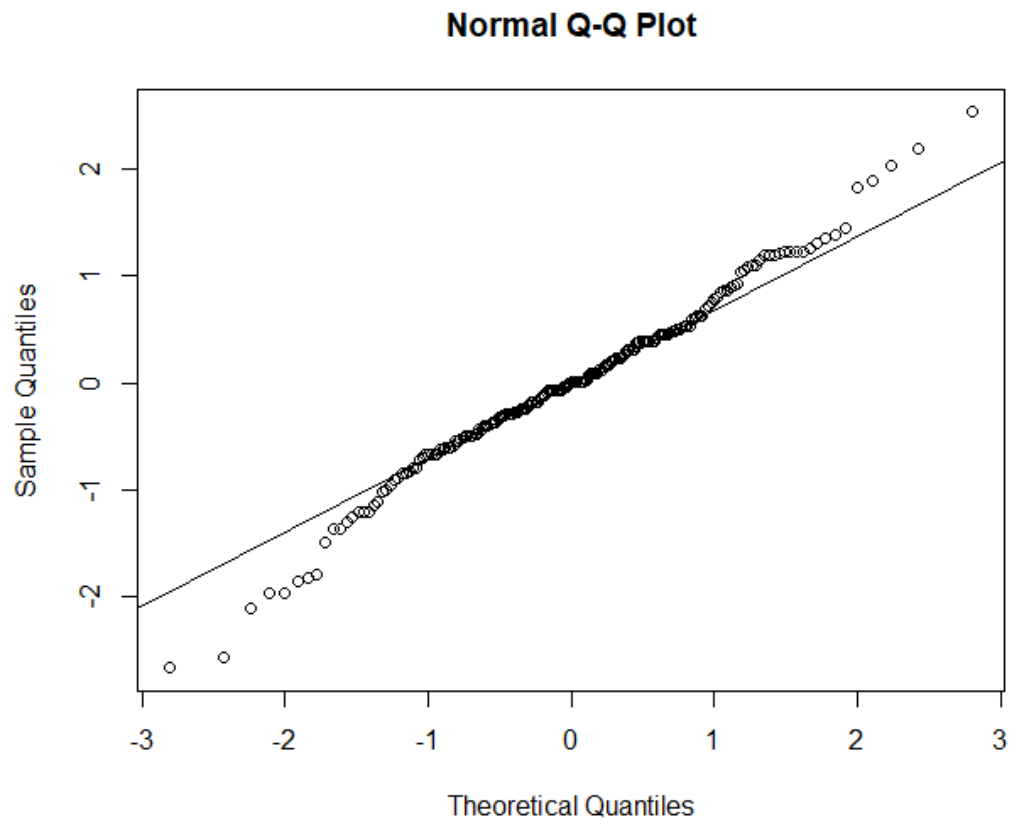
- Is dit actieplan gelijk voor alle leerlingen?
    - *Hoe sluit u aan op de leerbehoeftes van alle leerlingen?*
- Zou u tussentijds de progressie van leerlingen meten? Hoe zou u dit aanpakken?
- Betrekt u naast de leerlingen nog andere belanghebbenden in het behalen van de doelen? (Ouders, gemeenschap, collega's, enz.)

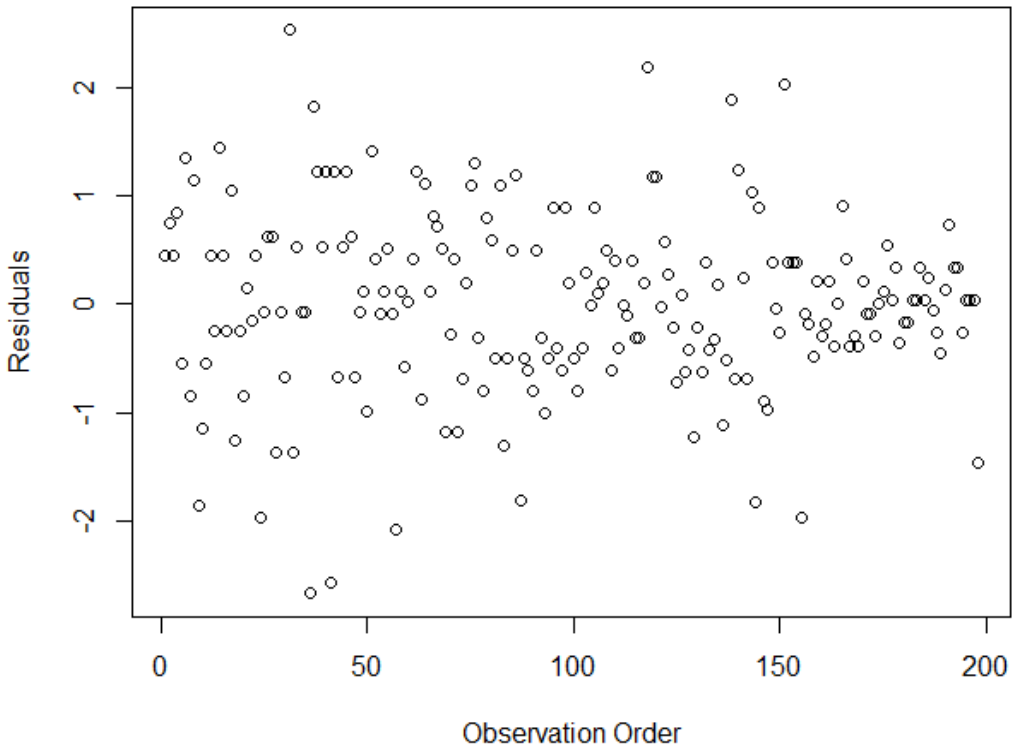
Leraren efficacy

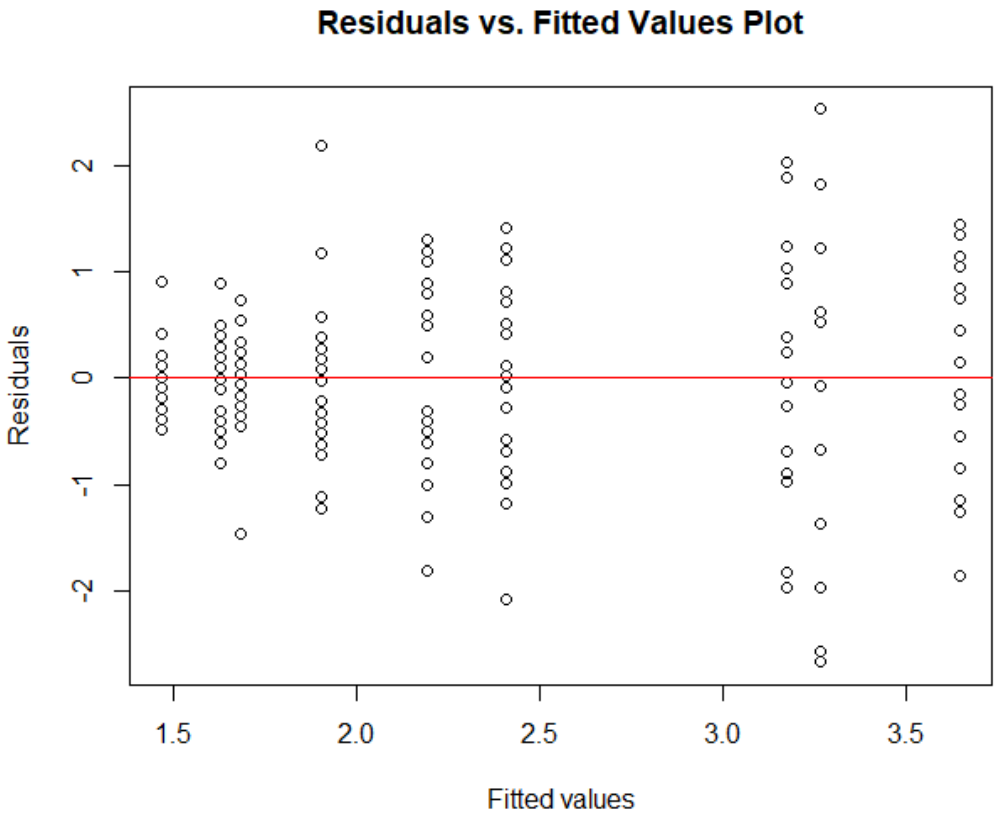
- In hoeverre bent u in staat om uw instructiemethoden aan te passen op basis van de behoeften van de leerlingen?
    - *Kunt u een voorbeeld geven uit de klas?*
- In hoeverre bent u in staat om een positieve en gestructureerde leeromgeving te creëren in de klas?
    - *Hoe realiseert u deze omgeving in de klas?*
- In hoeverre bent u in staat om leerlingen actief te betrekken bij het leerproces?

Kunt u een voorbeeld geven van een handeling die dit bevorderde

Appendix III: Plots






Residuals vs. Fitted Values Plot
Residuals
2
1
0
-1
-2
1.5
2.0
2.5
3.0
3.5
Fitted values

Appendix IV: R-code

```
library(tidyverse)
library(lme4)
library(lmerTest)

# Discriptive statistics
DiscriptivesDF <- summary(groupDF)
DiscriptivesGF <- Summary(groupGF)

SDDF <- sd(groupDF)
SDGF <- sd(groupGF)

# Perform an independent samples t-test
shapiro.test(groupDF)
shapiro.test(groupGF)

result <- t.test(TeacherDF, TeacherGF)
t_test_result <- t.test(studentDF, studentGF)

# Mixed-linear model
mixed.lmer2 <- lmer(Diff ~ Group + (1 | Teacher), data = dataset_2)
summary(mixed.lmer2)

bic <- BIC(mixed.lmer2)
qqnorm(residuals(mixed.lmer2))
qqline(residuals(mixed.lmer2))
plot(residuals(mixed.lmer2) ~ fitted(mixed.lmer2),
    xlab = "Fitted values", ylab = "Residuals",
    main = "Residuals vs. Fitted Values Plot")
abline(h = 0, col = "red")
```

```
plot(residuals(mixed.lmer2) ~ seq_along(residuals(mixed.lmer2)),
    xlab = "Observation Order", ylab = "Residuals",
    main = "Residuals vs. Observation Order Plot")

# Calculate the variance explained by fixed effects
lm_model <- lm(Diff ~ Group + test, data = dataset_2)
var_exp_lm <- 1 - (sum(residuals(lm_model)^2) / sum(residuals(mixed.lmer2)^2))
print(var_exp_lm)
```

Appendix V: Use of Generative Models

Generative models, particularly Chat-GPT 3.5 and 4.0, were used in the writing of the thesis. These models were used for two primary purposes: editing and idea generation. This appendix provides insight into how these models were used in the research and writing process.

Editing

The generative capabilities of Chat-GPT 3.5 and 4.0 were leveraged for text refinement purposes, particularly in the context of rewrites. Texts that had been written were presented to the models, accompanied by specific tasks such as shortening the text or restating arguments. Moreover, paragraphs were subjected to scrutiny for redundancy, with the models tasked to identify and suggest revisions. While single-sentence rewrites were occasionally adopted directly, more often, this process evolved into a collaborative effort, with suggestions from the models serving as co-writing prompts. Multiple rounds of refinement were conducted, with a selection of model-generated suggestions being applied iteratively.

Idea Generation

Generative models were employed as brainstorming tools to stimulate idea generation. Snippets of text were provided to the models, accompanied by specific questions aimed at exploring different perspectives or generating new insights. For instance, questions such as "What would be a counterargument for this statement?" or "Can you provide additional reasons for this phenomenon?" were posed to the models. Additionally, broader inquiries like "What can you tell me about [insert topic name]?" were utilized to generate starting points for research and ideas. It is crucial to note that the output generated by the models served as inspiration and prompts for further research rather than direct content incorporated into the thesis.

Appendix VI: Codebook

Topics	**Theme's**	**Codes**	**Data-first Group**		**Goal-first group**	
			Mentioned by the number of respondents (of 4)	*Examples (Respondent letter code A-E)* ***Participant D did not participate in the interviews***	*Mentioned by the number of respondents (of 4)*	*Examples (Respondent letter code F-I)*
Goals	Factors considered for goalsetting	Grades from study *Participant indicates the grades from the test provided for this study were used in the goal-setting procedure*	4	Participant C: "Neem bijvoorbeeld nummer 12, hij heeft een 4,5 maar ik weet dat hij normaal altijd goede cijfers haalt en hard studeert, veel vragen stelt. Weet je... echt een fijne leerling dan denk ik die kan echt wel een 7,5 halen." Participant A: "Ja ik heb eerste gesorteerd op het cijfer dat ze gehaald hadden en dan gekeken of dat voor mijn gevoel een beetje klopte"	0 (not possible)	
		Other Grades *Participant indicates student grades not resulting from this study were used in the goal-setting procedure*	3	Participant E: "Ja ik heb ook zeker gekeken naar de hun cijfers, als ze een 8 of een 9 staan dan ga ik er wel vanuit dat ze dat ook kunnen halen" Participant C: "Neem bijvoorbeeld nummer 12, hij heeft een 4,5 maar ik weet dat hij normaal altijd goede cijfers haalt en hard studeert, veel vragen stelt. Weet je... echt een fijne leerling dan denk ik die kan echt wel een 7,5 halen."	4	Participant F: "Neem bij voorbeeld XXX die staat echt wel goed op wiskunde maar ik weet gewoon die leert een week voor de toets alles, dus die doe ik dan een half puntje lager" Participant G: "Nee klopt... Heb de toetsen van dit jaar nog zeker even teruggekeken."

Intuition *Participants indicate some form of intuition or feeling was used in the goal-setting procedure*	2	Participant A: “Ja ik weet eigenlijk niet waarom ik dacht dat de een 7 zou halen en de ander een 8. Ze hebben wel hetzelfde op de toets. Dat is dan toch een onderbuikgevoel.” Participant B: “Het blijft toch ook wel een beetje een gevoelskwestie snap je”	3	Participant G: “Het kost me zo’n 15 seconden, ik dacht van waar staan ze, maar ook vooral hoe zijn ze in de klas, doen ze actief mee en maken ze hun huiswerk?” Participant I: “ja ik ben toch ook wel heel erg uitgegaan van mijn gevoel, ik ken die leerlingen en ik heb toch wel een idee wat ze kunnen “
Student study habits *Participant indicates the Study habits of students, including homework routines and test preparation practices, were used in the goal-setting procedure*	2	Participant C: “Neem bijvoorbeeld nummer 12, hij heeft een 4,5 maar ik weet dat hij normaal altijd goede cijfers haalt en hard studeert, veel vragen stelt. Weet je... echt een fijne leerling dan denk ik die kan echt wel een 7,5 halen.” Participant E: “even kijk hoor… die had een 5.2 op de toets, maar die haalt altijd 8en en 9ens die is heel serieus werkt zelfstandig voorruit daar heb ik geen omkijken naar vandaar die 8.5”	4	Participant F: “Neem bij voorbeeld XXX die staat echt wel goed op wiskunde maar ik weet gewoon die leert een week voor de toets alles, dus die doe ik dan een half puntje lager” Participant: "Ik heb echt wel opgelet, wie serieus met wiskunde bezig is, z'n huiswerk goed doet en netjes de leerlijn volgt."
Other student characteristics *Participants indicate student characteristics such as ability and personality not related to the Study*	2	Participant B: “Een van de meisjes, die is gewoon zo bang voor toetsen, doet het super in de klas maar die stort gewoon helemaal in bij een toets dus…. Daar hou je dan wel een beetje rekening mee”	3	Participant H: “Bij sommige zie je gewoon dat die er aanleg voor hebben, die pakken alles zo snel op. “

		habits of students, were used in the goal-setting procedure		Participant A: “Dat is zo eentje die gewoon niet wil… die zitten er omdat het moet”		Participant H: “Zij is heel veel afwezig, ik weet niet wat dat is maar ga ervan uit dat ze minder snel voorruit gaat”
Assessment of goal (attitudes)	Realistic/Attainable *Participant indicates the goal is achievable within the timeframe given normal circumstances*	4	Participant B: “dit moet gewoon beter, iedereen moet gewoon een voldoende kunnen halen.” Participant C: “Ze hebben dit allemaal al gehad, ik denk echt dat het goed te doen is”	3	Participant I: “Het is eigenlijk gewoon herhaling, ze hebben het allemaal al gehad” Participant F: “Ja, ik denk eigenlijk dat het wel realistisch is ja”
	Ambitious/Challenging *Participant indicates the goal requires significant effort and or dedication.*	2	Participant C: “Ik denk wel dat we hard aan de slag moeten.” Participant A: “Er zijn een aantal die gewoon een 1 of 2 scoren, misschien hebben ze niet hun best gedaan… maar denk dat daar wel echt veel moet gebeuren”	4	Participant F: “Ja 3 of 4 punten is wel veel, maar als ze hun best doen dan moet dat volgens mij gewoon kunnen” Participant G: “Gemiddeld ruim 3 punten is best een grote sprong. Maar het is allemaal bekend, ik heb er vertrouwen in ”
	Motivated *Participant indicates a measure of drive and intent to complete the goal.*	3	Participant E: “Ja ik vind dit soort dingen echt leuk, zoek graag de uitdaging op in mijn vak” Participant C: “Weet niet hoeveel tijd ik ervoor kan maken maar ik zie het op zich wel zitten”	4	Participant H: “Ik ben er wel van geschrokken, ik wil toch wel graag met mijn klas kijken hoe dit beter kan” Participant G: “Erg gemotiveerd, sinds ik de cijfers heb ben ik aan het denken hoe doen we dit anders.”
	Unmotivated	1	Participant A: “Heel eerlijk gezegd ben ik niet heel gemotiveerd om er	0	

		Participant indicates little to no drive and or intent to complete the goal.		mee bezig te gaan, het is echt wel interessant, maar de druk op leerlingen is al hoog genoeg"		
		Relevant *Participant indicates the goals and/or tests are relevant in the context of the current curriculum.*	4	Participant E: "Ja dit doet mij wel denken of we niet te veel voor de toets leren" Participant C: "Ik denk dat het zeker meerwaarde heeft, ik wil eigenlijk mijn andere klassen ook op deze manier gaan toetsen"	4	Participant I: "Ik vind het heel goed zo'n toets, drukt ons toch even met de neus op de feiten" Participant F: "Ik denk dat de toets heel goed aansluit op de leerlijn, het is allemaal stof voor het eindexamen"
		Uncertain *Participant indicates uncertainty about the goals.*	0		1	Participant F: "Ja met de kennis van nu had ik misschien wel een half puntje lager ingezet hier en daar"
Method assessment	Assessment of the Data-first method	Similar to normal practices *Participant indicates the method is similar to normal teacher practices.*		Participant E: "Kijk, het is net als je gewone gang van zaken, hè? Misschien niet zo hoogdravend, maar je hebt toch altijd een idee van waar iedereen staat, waar je naartoe wilt, en hoe je dat wilt bereiken. Het draait allemaal om doelen stellen en plannen maken" Participant B: "Het was eigenlijk best jammer, ik had echt graag bij die andere groep willen zijn. Dit voelt een beetje als 'been there, done that'. Maar vond ik de toets wel echt super interessant. Ik ben echt benieuwd of anderen klassen ook zo slecht hebben gescoord."	Not possible	
		Comprehensive *Participant indicates*		Participant C: "Het voordeel is dat je hier meer de tijd neemt, al was dit misschien wel wat veel. Normaal	Not possible	

			gesproken bedenk ik niet voor elke student een exact streefcijfer; soms laat ik hen dat wel zelf doen trouwens... Maar over het algemeen denk ik niet bewust na over wat iedereen zou moeten halen. Nou ja, ik bedoel, natuurlijk gebeurt het weleens dat je iets nakijkt en denkt, 'daar had ik meer van verwacht', maar dat is toch wel wat anders." Participant A: "Ja, dit was wel wat extra werk. Meestal hou ik wel in de gaten hoe de klas ervoor staat, maar zelf stel ik meestal geen specifieke doelen voor toetsen. Natuurlijk hoop ik dat iedereen het haalt, maar uiteindelijk ligt dat ook een beetje in handen van de leerlingen zelf, toch?"		
	Other		Participant C: "Weet je, wat me echt aansprak, is dat je nu meer aandacht hebt voor de mensen in het midden. Vaak zijn we gefocust op de achterblijvers en de toppers, maar die middenmoot wordt soms vergeten. Nu denken we echt na over wat zij kunnen bereiken en dat vind ik geweldig."	Not possible	
Assessment of the Goal-first method	Complex *Participant indicates that setting goals using this method is harder than normal practices*	Not possible			Participant F: "Moet zeggen, ik vond het behoorlijk pittig. Die toets was echt heel anders dan wat ze normaal gewend zijn, met veel meer onderwerpen. En ze hadden

						natuurlijk ook niet van tevoren geoefend, plus moesten ze het zonder rekenmachine doen."
		Personal *Participant indicates that the method makes the user consider more factors related to the student's personality and habits.*	Not possible			Participant G: "Ik vond deze manier van werken persoonlijk erg interessant, maar ook een beetje vreemd. Ik moest omschakelen. Het dwingt je om anders te denken, meer aandacht te besteden aan wie iemand is, hoe ze zich gedragen... Ik vind het persoonlijker." Participant I: “Deze methode trok echt mijn aandacht; het liet me meer nadenken over de kinderen zelf, niet alleen over cijfers. Het draaide echt om wie ze zijn en wat ze nodig hebben in de klas”
		Challenged preconceptions *Participants indicated the method confronted wrong beliefs or estimates.*	Not possible			Participant H: "Ja, het is echt gek om je gevoel zo op de proef te stellen, helemaal als blijkt dat je dan ook nog eens flink mis zit. Maar eerlijk gezegd, het is wel echt goed. Ik wist natuurlijk wel dat leerlingen dingen na verloop van tijd vergeten, maar dit was toch wel op een ander niveau dan ik had gedacht. Het was even schrikken, maar ja nu weet ik het en kan ik er iets mee."
Action plans development (WHY/How)	Formal and informal data used	Test data (from the study) *Participant indicates data from the test provided for this study, such as common*	2	Participant A: “Ik snap dat niet, het bepalen van het snijpunt hebben we laatst nog behandeld en toch heeft bijna de helft het fout… Nou daar	2	Participant G: “Ik wil dat op doen op basis van de toets bespreking, dan kunnen de leerlingen zelf aangeven wat zelf wel en niet lukt”

	mistakes, was or will be used in the development of the action plan		moeten we dan in ieder geval meer tijd aan besteden”		Participant I: “Ik heb de toets niet voor mij, maar ik zie in mijn bestandje dat vraag 2,5,7 en 8 het slechts gemaakt zijn daarmee zou ik beginnen”
	Other assessment data *Participant indicates data from a previous or future test was or will be used in the development of the action plan (except data from the test associated with this study)*	3	Participant C: “Ik ben wel even aan het gelijken geweest met oude toetsen, vereenvoudigen ging nu bijvoorbeeld heel moeilijk maar 2 maanden geleden ging dat nog veel beter, daarbij zou ik dan denken dat ze het zelf wel kunnen herhalen…”	2	Participant H: “Met dit soort dingen moet je tussentijds ook een beetje schakelen, ik zou denk ik wel 1 of 2 toets moment tussendoor doen gewoon even peilen waar iedereen is”
	Input from students *Participant indicates student input, among which verbal or written utterances from students except test answers, were or will be used in the development of the action plan.* ***** If any additional information about a test, which is not a direct answer to a graded question, is provided it counts as student input.*	2	Participant B: “Ik plan met dit soort dingen graag niet te gedetailleerd, ik vind je moet de leerbehoeftes in de klas bekijken en dat meenemen in hoeveel tijd je aan onderwerpen besteed” Participant E: “Als ik een toets afneem vraag ik altijd om de moeilijkste vraag te omcirkelen, nu was dat vraag 9b…het berekenen van die halve cirkel”	3	Participant G: “Ik wil dat op doen op basis van de toets bespreking, dan kunnen de leerlingen zelf aangeven wat zelf wel en niet lukt” Participant G: “… zet ik de leerlingen na de toets in 2 of 3 groepen en dan moet elke groep overleggen welke vragen het moeilijkst waren, dan doen we die nog een keer samen op het bord”
	Experience *Participant indicates previous experiences were used to develop the action plan*	4	Participant E: “Ik doe dit al 20 jaar, op een bepaald moment weet je wel hoe je een lesplan maakt” Participant B: “Poeh waarom doe ik dat zo? ...Daar vraag je mij wat….	4	Participant I: “Oh, ik had vroeger een leraar die altijd op de toetsen vroeg naar de kleuren van de regenboog, en dat is me altijd bijgebleven, snap je? Dus ik dacht, waarom zou ik niet iets soortgelijks

			Ik denk dat het ook een stukje ervaring is.”		proberen? Oude onderwerpen terugbrengen in de toetsen, zodat studenten blijven herhalen en onthouden wat ze hebben geleerd.” Participant F: “Ja door de tijd heen probeer je het een en ander, ik merk dat voor veel leerlingen gewoon heel fijn om tijdens de les klassikaal opdrachten te maken samen dat begrip op te bouwen”
	Scientific literature *Participant indicates scientific data was or will be used in the development of the action plan*	1	Participant E: “Ik ben wel bekend met het Pygmalion effect, ik probeer altijd voorzichtig te zijn met mijn verwachtingen”	0	
Constraints	Time *Participant indicates available time is a limiting factor in the action plan.*	3	Participant B: “Met 2 lessen per week zie ik gewoon beperkt de mogelijkheid om hiermee aan de slag te gaan” Participant A: “Ik weet niet waar ik de tijd vandaan moet halen, ik ben laatst ziek geweest dus we lopen al achter…. Misschien aan het eind van het jaar maar dat moet ik echt dan bekijken”	1	Participant H: “Weet je wat het is, de lessen zitten gewoon al heel vol. Ik kan zulke dingen gewoon niet oppakken zoals ik dat graag zou willen. “
	Materials *Participants ask for additional materials.*	1	Participant B: “Waar komt die toets eigenlijk vandaan? Heb jij die zelf gemaakt?”	3	Participant F: “Ik had nog een vraag, heb jij nog een toets die ik eind van het jaar kan inzetten?” Participant I: “Je had het in de mail ook nog over een toets voor klas 3 kan ik die ook krijgen?”

	Other *Participant indicates any limitation that cannot be categorised in any of the previous categories.*	2	Participant C: “Ik en een meester van de andere klas volgen hetzelfde lesplan, we overleggen en bespreken elke paar weer de voortgang even. Hij ziet zo het niet zitten om dingen aan te passen” Participant A: “Ik weet gewoon echt niet waar ik moet beginnen, ze maken fouten op dingen die we laatst behandeld hebben”	1	Participant H: “Ik wil hier wel wat mee maar het is best lastig… Mijn leerlingen hebben het al moeilijk genoeg”
Intended outcomes	Mindset change/behaviour change *Participant indicates behavioural change, such as study behaviour and behaviour in class is an intended outcome.*	1	Participant C: “Het liefst wil je natuurlijk dat iedereen zijn huiswerk netjes maakt en een beetje serieus meedoet met de lessen”	3	Participant F: Het moet echt doordringen dat je niet even weekje van tevoren kan leren voor zo’n toets, dan komt het gewoon niet in de lange termijn. Ze moeten bijhouden, bijhouden, bijhouden. Participant H: Het is en blijft lastig om de study gewoonten van leerlingen te veranderen, maar ik ga het toch nog een keer proberen, misschien dat deze toets wel de aanwijzing is die ze nodig hebben. Echt aan ze laten zien dat het zonde is dat dingen niet goed beklijven…. Denk dat die oude vragen toevoegen tevens goede motivatie voor ze zal zijn.
	Improved grades/academic improvement	4	Participant E: “Het is eindexamen stof, dus ja hier moeten ze wel mee bezig eind van het jaar moeten ze	4	Participant G: “Iedereen moet gewoon een voldoende kunnen halen, dus daar gaan we ook voor!”

		Participant indicates higher grades as an intended outcome.		echt wel een paar punten beter scoren" Participant A: "Eind van het jaar wil ik gewoon iedereen minimaal 2 punten hoger zien scoren"		Participant F: "Natuurlijk zijn cijfers ook van belang, hier zie je uiteindelijk toch terug of het allemaal een beetje gewerkt heeft"
Action plan content (What)	Curriculum *Action plan elements related to non-graded assignments and instructions.*	Modifying the content of existing lessons. *Participant indicates elements of the action plan are integrated into changes in existing materials and curricula*	1	Participant B: "Ik dacht eraan om het via het huiswerk te regelen. Gewoon elke week een opdracht toevoegen van een vorig hoofdstuk. Zo kunnen we tijdens de huiswerkbespreking dat onderwerp ook nog even doornemen als dat nodig is."	3	Participant G: "Eerst wil ik samen met de studenten de toets even doornemen…… maar wat mij wel wat leek, elke week een oude opdracht dacht misschien uit oude toets, op het bord samen uit te werken" Participant I: "Ik wil in mijn lessen de link met oud materiaal sterker leggen… opdrachten uitzoeken waarbij meerdere onderwerpen voorkomen"
		Separate lesson *Participant indicates separate lessons are organised in their entirety for instructions or assignments directly related to achieving the goal ability level.*	2	Participant E: "In de planning heb ik 2 of 3 lessen voor als ik ziek word of iets maar ik dacht wat als ik die nou inzet om de stof nog een keer in vogelvlucht te doorlopen, moet ik natuurlijk niet ziek worden (lach)" Participant C: "Eind van het jaar heb ik meestal wel extra lessen over. Als dat dit jaar weer het geval is, overweeg ik om één les te gebruiken voor vragen en de ander voor de toets"	2	Participant F: "Volgende week heb ik al een les vrijgemaakt, wil de toets even samen door, helemaal die vragen waar bijna de klas de fout in is gegaan. 3 en 9 zo uit mijn hoofd" Participant G: "Zou dan toch wat aan het schuiven moeten kijken of ik wat lessen vrij kan maken" Participant I: "Als ik die toets voor klas 3 nog van jou kan krijgen dan wilde ik die aan het eind van het jaar doen als check"

	Behavioural intervention *Participant indicates assignments, instruction or checks are implemented with the intent of changing/facilitating student behaviour, such as study habits and behaviour in class.*	2	Participant C: “Nou ik wil toch weer strenger huiswerk gaan controleren, vind dat eigenlijk altijd wat betuttelend, maar je ziet nu maar weer hoe belangrijk het is dat ze bij blijven” Participant E: “We zijn in overleg of wij tijdens de steruren (uitvaluren) vaker bezig kunnen met een soort huiswerkbegeleiding voor wiskunde en NASK”	3	Participant I: “Het is en blijft lastig om de study gewoonten van leerlingen te veranderen, maar ik ga het toch nog een keer proberen, misschien dat deze toets wel de aanwijzing is die nodig is om echt aan ze laten zien dat het zonde is dat dingen niet goed beklijven.” Participant G: “Ik overweeg om de komende maanden s’avonds een paar keer videogesprekken te organiseren waar dan samen met mij het huiswerk kunnen maken”
	Repeating instruction and assignments. *Participant indicates previous materials are repeated for practice (non-graded) as part of the action plan.*	3	Participant B: "Ik dacht eraan om het via het huiswerk te regelen. Gewoon elke week een opdracht toevoegen van een vorig hoofdstuk. Zo kunnen we tijdens de huiswerkbespreking dat onderwerp ook nog even doornemen als dat nodig is."	3	Participant G: ‘Eerst wil ik samen met de studenten de toets even doornemen…… maar wat mij wel wat leek, elke week een oude opdracht dacht misschien uit oude toets, op het bord samen uit te werken’
	Additional Homework *Participant indicates additional homework assignments will be implemented to achieve the goals*	2	Participant B: "Ik dacht eraan om het via het huiswerk te regelen. Gewoon elke week een opdracht toevoegen van een vorig hoofdstuk. Zo kunnen we tijdens de huiswerkbespreking dat onderwerp ook nog even doornemen als dat nodig is."	1	Participant F: “Extra huiswerk is ook zeker een optie, maar ik wil eerst even kijken of het binnen de lessen past.”
	No change *Participant indicates no intention to change*	1	Participant A: “Ik zie geen mogelijkheid om er extra tijd aan te	0	

	anything related to the assignment or instruction.		besteden, dat moet maar in de eigen tijd”		
Assessment	Modify existing assessment. *Participant indicates intended changes to previously planned assessments related to achieving the goal.*	1	Participant C: “Een bonusvraag aan het einde over dingen wat ooit al een keer behandeld”	3	Participant I: “Oh, ik had vroeger een leraar die altijd op de toetsen vroeg naar de kleuren van de regenboog, en dat is me altijd bijgebleven, snap je? Dus ik dacht, waarom zou ik niet iets soortgelijks proberen?” Participant H: “Toetsen zouden meer 20 80 structuur moeten krijgen, 20 procent uit eerdere periodes en 80 procent uit het huidige hoofdstuk”
	Additional assessments *Participant indicates additional assessment is added related to the set goal.*	3	Participant B: “Ik overweeg om eind van het jaar weer zo’n toets opnieuw te doen, maar deze keer van tevoren de onderwerpen duidelijk met de leerlingen bespreken en wel echt voor een cijfer” Participant E: "Denk dat het best een goed idee is om aan het eind van het jaar zo'n soort toets te hebben, een soort van examen, waarin alle onderwerpen van dit jaar kunnen zitten."	2	Participant F: “Ik had nog een vraag, heb jij nog een toets die ik eind van het jaar kan inzetten?” Participant G: “Mijn intentie is wel om blok 4 de toets weer te doen, ik wil toch wel weten of er straks nou echt iets veranderd is”
	Other *Participant indicates any changes to the assessment related to achieving the set goal which does not fit in*	0		1	Participant H: “Ik heb de andere leraren nog niet mee, maar ik wil aan het eind graag dat ze reflecteren over hoe ze hebben geleerd, of ze het huiswerk hebben gemaakt, en of

	any of the previous categories.				ze iets anders zouden doen… Voor een kleine bonus.”
Involving the environment	Involving parents *Participant indicates a willingness to involve parents in achieving action plans.*	1	Participant B: “Soms neem je de omgeving wel mee natuurlijk dan zet je bij de ouderavond de boel toch even op scherp”	0	
	Involving other teachers *Participant indicates a willingness to involve other teachers in achieving action plans.*	1	Participant C: “verder trek je natuurlijk ook welleens bij een mentor aan de bel als het echt niet lekker loopt”	0	
Differentiation (between students)	Individual *Participant indicates differentiation at an individual level*	2	Participant E: “In deze klas zitten nu 3 studenten die hun eigengang gaan, er waren 4 maar bij eentje liep het niet helemaal goed, die haalde oppeens een 6 dus ja dan moet die eerst maar weer even met de rest meedoen”	2	Participant G: ”Probeer altijd zo veel mogelijk op individuele leerbehoeften aan te sluiten, sommige kinderen gaan vrij aan de slag, bepalen soms helemaal zelf wat ze een les doen, die snappen het zelf wel, maar anderen ga je bijna elke les mee zitten samen opdrachten maken.”
	Groups: same level *Participant indicates differentiation at the group level, using participants with similar abilities.*	2	Participant B: Ik moedig leerlingen wel aan om samen aan de opdrachten te werken, maar dan moet het niet zo zijn dat eentje meelift natuurlijk	1	Participant F: “Vaak geef ik eerst klassikaal uitleg en dan kan iedereen die het snapt aan de gang en wie nog wat extra uitleg kan vooraan samen de eerste opdracht met mij maken”
	Groups: mixing levels *Participant indicates differentiation at the group level, using participants with different abilities.*	1	Participant A: “Er zijn een aantal die het heel makkelijk snappen soms vraag ik een van hun weleens om dan samen te werken met iemand die het even moeilijk vindt… denk	2	Participant G: “Als een onderwerp echt helemaal nieuw is dan begin ik graag in groepjes, dan probeer ik de beste van de klas een beetje gelijk te verdelen”

				dat ze daar beide wel van leren uitleggen kan ook heel leerzaam zijn”		
Teacher efficacy	Instructional strategies	High *Evident when educators possess a strong self-perception of confidence and effectiveness in employing a variety of instructional strategies. High efficacy in instructional strategies is characterized by a positive self-image, a belief in the impact of diverse teaching methods, and a proactive approach to adapting strategies for optimal student learning.*	2	Participant E: “Oh dat is iets waar wij als team heel veel mee bezig zijn, we weten hoe lastig wiskunde voor sommige is dus ja daar doen we heel veel mee, dus ja daar weet ik wel raad mee”	2	Participant G: “Voor mij is het echt belangrijk om zoveel mogelijk op maat les te geven. Vooral bij wiskunde raken sommige leerlingen snel ontmoedigd als iets niet lukt. Het maakt een groot verschil als je goed kunt aansluiten op hun niveau, zodat ze succeservaring kunnen opdoen, snap je?"
		Medium *Signifies a self-perception of moderate confidence and competence in implementing instructional strategies. Educators at this level may acknowledge the need for improvement in certain areas but generally maintain a positive self-perception regarding their ability to use a range of effective teaching methods.*	2	Participant A: “Op zich maar ik ben van origine geen wiskunde leraar denk dat ik in de opleiding nog wel het een en ander zou kunnen leren”	2	Participant I: “Meestal lukt dat wel aardig maar soms is het best uitdagend als iemand er heel negatief instaat”
	Classroom management	High	4	Participant B: “Vroeger vond ik dat soms best ingewikkeld, maar na 15	3	Participant H: “Ik gebruik ook veel positieve bekrachtiging,

		Evident when educators possess a strong self-perception in effectively navigating and controlling classroom dynamics. High efficacy in classroom management includes a positive self-image, the ability to establish clear expectations, and confident handling of incidents.		jaar voor de klas te hebben gestaan, leer je dat wel… Het zijn pubers, dus er is echt af en toe weleens een incidentje, maar ik durf wel te zeggen dat de sfeer eigenlijk altijd wel goed is." Participant C: "Nou, ik begin elk schooljaar met het samen opstellen van klassenregels. Zo voelt iedereen zich betrokken. "		complimentjes voor goed werk dit soort dingen, en probeer altijd tijd te maken om naar de leerlingen te luisteren en hen te helpen waar nodig"
		Medium *Reflects a self-perception of moderate efficacy in managing classroom behaviour. Educators at this level may acknowledge variability in their ability to maintain order but generally maintain a positive self-perception regarding their competence in handling classroom dynamics.*	0		1	Participant I: "hahaha die vraag hangt voor mij echt af van de klas die ik heb, begrijp mij niet verkeert, ik kan wel orde houden hoor maar er zijn van die klassen waar het af en toe gewoon chaos is"
	Student engagement	High *Demonstrated when educators actively and confidently perceive their creative and effective strategies to engage students. High efficacy in student engagement involves a positive self-*	2	Participant E: "Ik merk dat gewoon laten zien hoe iets werkt vaak veel beter is dan het uitleggen. Als ik vroeger op de basisschool conversies uitlegde, nam ik altijd mijn doos van 10 bij 10 mee. Je weet wel, dat een liter een kubieke decimeter is, toch? Dan vroeg ik altijd wie wil wedden dat er een liter	1	Participant G: "Ja ik vind dat ik dat wel kan ja, toch altijd wel het gevoel dat de meeste kinderen wel actief bij mij in de les zitten, al blijft het natuurlijk wiskunde he"

	perception of fostering an interactive and participatory learning atmosphere.		in past. Gewoon kleine dingen die tot de verbeelding spreken."		
	Medium *Indicates a self-perception of moderate efficacy in engaging students. Educators at this level may find certain aspects of student engagement challenging but generally perceive themselves as striving to create an inclusive and participatory learning environment.*	2	Participant A: “Soms vind ik dat best uitdagend, ik bedoel hoe zou jij wiskunde aantrekkelijk maken voor een stel pubers?”	3	Participant F: “Zelf richt ik mij toch echt wel op de leerlingen die mee willen doen, zie dat toch wel als eigenverantwoording”