Tootling for Supporting the Inclusion of Students with Special Needs in General Education Classrooms: A Systematic Literature Review.

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Applied Psychology

(Behaviour Analysis)

at

The University of Waikato

by

Alevtina Kabalinskaya

Student ID 1559889

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2025

Abstract

Disruptive behaviours negatively impact students' academic and social progress, leading to poor outcomes later in life. The effects can be especially detrimental for students with additional needs. After decades of reliance on a punitive approach to behaviour management in schools, positive behaviour interventions are gaining recognition for their ability to manage problematic behaviours by teaching students the appropriate ways to interact and creating positive and supportive environments. The Schoolwide Positive Behaviour Interventions and Support (SWPBIS) framework provides a proactive, flexible approach to behaviour management through a tiered system of supports. Supports are provided in accordance with the needs of a student, with the strategies of Tier 1 designed to support the needs of the majority of students. Considering that physical placement of special needs students in general education classrooms does not always mean meaningful inclusion of those students on academic and social processes of the classrooms, strategies for inclusion are needed. As inclusion strategies should target both typical and special needs students, they should be implemented schoolwide, as Tier 1 strategies. In this systematic literature review a positive behaviour intervention called Tootling is evaluated as a potential strategy for enhancing inclusion of students with special needs in general education classrooms. Being implemented classwide Tootling has a potential improving behaviour of special needs students without separating or singling them out in any other way.

The review examined sixteen studies that assessed the effects of Tootling at the classwide level in inclusive classrooms, alongside three studies that focused on individual students with special needs. Studies were evaluated using the What Works Clearinghouse (WWC) standards for research design quality (Institute of Education Studies, 2017). Data were systematically aggregated to provide an overview of participants, methodologies, and

findings. Additionally, the outcome data was subjected to systematic visual analysis for additional validation.

Findings indicate that Tootling is an effective, evidence-based intervention for reducing disruptive behaviour and increasing appropriate behaviour in inclusive classrooms, meeting the WWC “5-3-20” threshold (Institute of Education Studies, 2017). The evidence for the effect of Tootling on passive off-task behaviour was insufficient with a small number of studies that evaluated passive off-task and inconclusive results of those studies. The results regarding the effect of Tootling on disruptive and appropriate behaviour in individual students with special needs were promising. Further research from various research teams covering larger numbers of diverse participants is needed to strengthen the evidence base and explore long-term impacts.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Angelika Anderson, for her patience, support, and wisdom throughout this journey. Her guidance has been invaluable, and I am truly grateful for her encouragement and insight.

To my husband, thank you for your unwavering support over these five years of study. Your belief in me kept me going through every challenge.

To my daughter, I am incredibly grateful for your understanding and support. Your patience and kindness made this journey easier.

A heartfelt thank you to my mother and my mother-in-law for their incredible support and care of our family and especially their little grandson, who was born only four months before I started my postgraduate studies. Your love and dedication allowed me to focus on my studies, knowing he was in the best hands.

And to my son - thank you for being such a good boy and allowing whole family to care for you. Your little smiles and hugs have been a constant source of joy and motivation.

This thesis is a reflection of not just my effort, but the guidance of my supervisor and the support and love of my family. I am forever grateful.

Table of Contents

List of Figures

List of Tables

Literature Review

Importance of Classroom Behaviour Management

Disruptive, or problematic, student behaviours are actions that hinder the social, emotional, and academic development of students and interfere with teachers' ability to conduct their work effectively (Artman-Meeker & Hemmeter, 2013; Wheldall & Merrett, 1988). These behaviours may include inappropriate vocalisations, being out of seat without permission, and engaging in activities unrelated to the task at hand (Lambert et al., 2015). Such behaviours are prevalent issues in classrooms worldwide (Browne, 2013) and in New Zealand (Towl, 2013).

The majority of teachers consistently report feeling unprepared to manage these behavioural issues (Johansen et al., 2011) and that student misbehaviour interferes with their ability to teach effectively (Robers et al., 2014). The inability to address behavioural issues reduces the time available for academic activities, subsequently diminishing academic achievement (Lassen et al., 2006; Riley et al., 2011), as well as affecting behavioural and social development (Lynne Lane et al., 2007; Närhi et al., 2017). In environments with high levels of problem behaviours, even initially well-behaved students may begin to engage in negative behaviours (Barth et al., 2004). Disruptive students often take up a large portion of teachers' time and attention, affecting not only the disruptive students but also their peers (Walker et al., 2004).

Disruptive behaviours in the classroom have long-term consequences for the students concerned. These behaviours can lead to future difficulties, including mental health issues, criminal activity, drug use, and relationship problems (Bradley et al., 2008; Church, 2003; Greer-Chase et al., 2002; Suldo et al., 2014) later in life.

Environments with high rates of problem behaviours are not conducive to acquiring effective social skills. Students with disabilities and those from low socioeconomic status

(SES) backgrounds are more likely to experience social skill deficits, making it harder for them to establish positive relationships with peers (Gresham, 2002; Ng et al., 2018) and making them more likely to face repeated rejection in the future (Coie & Cillessen, 1993). These social skills deficits limit their opportunities to learn and increase the risk of social isolation and poor mental health (Henricsson & Rydell, 2006; Segrin & Flora, 2000). Therefore, disruptive behaviours have detrimental effects on students affecting their academic and social life and leading to poor outcomes later in life. The effects are especially detrimental for students with additional needs.

Punitive Strategies

Historically, behaviour management in schools centred on the enforcement of compliance through punishment of inappropriate behaviours (Henington & Skinner, 1998). Punishment usually presented independent group contingencies, meaning that only the student believed to commit rule-breaking suffers consequences and the same punishment is delivered for the same inappropriate behaviour across the group (Hayes, 1976; Litoe & Pumroy, 1975). Such contingency ensures that the student experiences consequences of their own inappropriate behaviour, promoting personal responsibility, and the punishment is fair for all (Skinner et al., 1999). While this approach is widespread and often reduces the instances of certain inappropriate behaviours, it comes with serious negative side effects.

First, this system places a significant burden on teachers, requiring constant vigilance and attention. Teachers must monitor the behaviours of all students, issue reprimands, and administer punishments. Given that observing all instances of problem behaviours is impossible, teachers often have to rely on peers reporting their classmates' inappropriate behaviours, a practice commonly known as tattling (Skinner, 2002). This reliance on tattling results in teachers spending excessive time and energy investigating these negative peer reports. Research shows that responding to tattling leads to loss of academic instructional

time (Collins et al., 2020). Additionally, teachers might make mistakes or reach incorrect conclusions, punishing innocent students or failing to punish guilty ones. Such mistakes may cause students and their parents to distrust their teachers (Henington & Skinner, 1998).

The punitive approach works differently on different students (Iwata et al., 1994). While some students avoid punishment by following the rules, others might learn to avoid punishment by not getting caught, developing maladaptive behavioural strategies (Skinner et al., 2000). Moreover, such systems can inadvertently encourage more problem behaviours. For example, a student who exhibits aggressive behaviour towards another student may threaten further aggression if the incident is reported, thereby deterring victims from speaking up and perpetuating a cycle of misbehaviour (Skinner et al., 2000). Research indicates that when such approaches are implemented with students with established histories of problem behaviours, they are likely to increase the severity and frequency of those behaviours (Sugai & Horner, 2002). Therefore, such approaches further aggravate problems in students with emotional and behavioural difficulties.

A significant drawback of the punitive system is that it places a constant focus on inappropriate behaviours, both for teachers and students (Skinner et al., 2000). Research shows that teachers spend more time focused on inappropriate than on appropriate behaviours (Thomas et al., 1978; White, 1975). Such focus means that appropriate behaviours often remain unnoticed and unrewarded, creating an unbalanced environment with limited opportunities for the development of positive behaviours. By concentrating primarily on what students should not do, the system fails to promote and reward what they should do, hindering the overall behavioural and social development of students. Students, by focusing on inappropriate behaviours, fail to notice their peers' pro-social behaviours and develop unbalanced perceptions of their classmates (Henington & Skinner, 1998).

Often, following severe cases of misconduct, schools increase reactive measures. Research demonstrated that the effects of these measures are short-lived. These measures give a false sense of appropriate action whereas in fact they do not improve the school environment. They establish authoritarian control and inadvertently reinforce anti-social behaviour (Sugai & Horner, 2002). Additionally, such reactive approaches, where action is delivered after the problem behaviour occurred, do not provide means to communicate and establish repertoires of appropriate, pro-social behaviours (Sugai & Horner, 2002). Students in these environments have little opportunities to develop appropriate behaviours. This is the so-called discrepancy or wait-to-fail model. If an individual student repeatedly fails, he or she receives a diagnosis of emotional/behavioural disorder or a learning disability. This approach suggests that there is something inherently wrong with the student without looking into the environment.

Overall, punitive strategies place all responsibility on students and enforce compliance through punishing inappropriate behaviours. While they may reduce some misbehaviours, they burden teachers, foster distrust, and often exacerbate issues for students with emotional and behavioural difficulties. Reactive measures do not address root causes, failing to provide a pathway for the development of appropriate behaviours.

Applied Behaviour Analysis

Applied Behaviour Analysis (ABA) is an applied branch of the scientific discipline behaviour analysis, which focuses on understanding and influencing animal and human behaviour through systematic methods of natural science. Emerging from research conducted in the mid-20th century, ABA has evolved into a robust framework for working with socially significant behaviours. ABA supports individuals in acquiring new skills, reducing problematic behaviours, and improving overall quality of life, particularly in educational and therapeutic settings.

The guiding principle of ABA is that behaviour is a function of the environment and can be altered by manipulating environmental factors. ABA examines behaviours by analysing their antecedents and consequences. This framework known as a three-term, or ABC, contingency with A for antecedents, B for behaviour and C for consequences. In this framework altering either antecedents (A) or consequences (C) leads to changes in behaviour (B). Inevitably, human behaviour occurs in complex environments, influenced by multiple antecedents and extensive histories of consequences. Nonetheless, research demonstrates that modifying relevant environmental factors can effectively lead to behaviour change.

Behaviours that are met with favourable outcomes tend to increase, while those that do not, tend to diminish. Reinforcement, defined functionally rather than by intent, increases the likelihood of a behaviour. For instance, additional playground time might reinforce a child's behaviour if that child enjoys outdoor activities but would have no effect on one who prefers indoors. Similarly, punishment reduces behaviour and is also defined functionally in Behaviour Analysis. The colloquial understanding of punishment is the use of an aversive stimulus to penalise a person for their behaviour, with the expectation that this will deter them from repeating their problematic action. However, aversive stimuli introduced with the intention to punish - such as in independent group contingencies in schools, where the same aversive strategy is applied to all students - may not actually decrease behaviour and, therefore, may fail to function as punishment. For example, scolding may appear like a punishment, but if a child seeking attention misbehaves more after being scolded, the scolding serves as reinforcement instead. This difference is one of the reasons why intended punishing procedures may not work in schools.

Punishment, as defined in ABA, may appear to be a logical solution for addressing problem behaviours. However, extensive research established that its use comes with significant risks (Cooper et al., 2020). The main learning produced by punishing procedures

is avoidance. Punishment does not provide opportunities for establishing a functional behaviour repertoire, but it can model undesirable behaviour leading to aggravation of behaviour problems. Other side effects include eliciting undesirable emotional responses and behavioural contrast: increased problem behaviours under non-punishment conditions. Punishment may be overused due to its immediate effects, which function as a reinforcement for the persons who punishes. Therefore, punishment may exacerbate anti-social behaviours in both children and their caregivers leading to a viscous cycle of dysfunctional behaviours. In order to be effective, punishment must be applied consistently and immediately after the problem behaviour (Cooper et al., 2020). This can be challenging in school settings, where teachers may not always be aware of students' inappropriate actions. Additionally, students often attempt to conceal such behaviour to avoid aversive consequences, further limiting the effectiveness of punishment.

Given these concerns, modern standards include a range of safeguarding strategies. Thus, according to the Statement on Restraint and Seclusion from the Association of Behaviour Analysis (Vollmer et al., 2011) the least restrictive alternative has to be considered first and punishment is used as a last resort strategy when all other strategies are exhausted and the benefit of punishment overweighs it’s potential risks. For example, punishment may be beneficial in the cases of severe self-injurious behaviour (Lerman et al., 1997). Otherwise, strategies such as antecedent modifications, differential reinforcement or extinction are used. As an example of antecedent intervention, a child who hits their caregiver to ask for something can be taught appropriate ways to ask. Such antecedent measures reduce the need for hitting while fostering social skills. Further, caregivers or therapists interact with a child in such a way that hitting behaviour does not result in satisfying the child’s need while asking does. Such interactions illustrate the implementations of differential reinforcement with extinction: appropriate behaviour is reinforced while inappropriate behaviour is put to

extinction and does not result in any reinforcement. Although extinction can produce side effects like emotional outbursts or increased problem behaviour, combining it with differential reinforcement mitigates these effects. Teaching alternative behaviours and differentially reinforcing them not only reduces undesirable behaviours but also helps learners acquire socially appropriate skills to achieve their goals in the future. Functional behaviour assessment (FBA) helps to identify the reinforcement maintaining the behaviour in question and design a proper intervention (Sigafoos et al., 2021).

In conclusion, ABA offers a comprehensive framework for understanding and influencing behaviour. Its principles, grounded in decades of research, provide educators with effective tools for managing classroom behaviours. By prioritising reinforcement over punishment and focusing on teaching new behaviours, ABA strategies align with the ethical imperative to ensure safe, humane, and effective interventions in educational settings. Additional focus on the rights of students and the need for consistency in applying these strategies further underscores the importance of a whole-school framework for behaviour management.

Inclusion

Inclusion in education is a process aimed at breaking down barriers that limit learners' presence, participation, and achievement (UNESCO, 2017). Historically, students with special needs were educated in special schools or classrooms, separately from their neurotypical peers. This changed when the Salamanca Statement on Inclusion in Education was adopted in 1994, establishing that all children should learn together in schools, regardless of differences or difficulties. In New Zealand, the policy of mainstreaming students with special needs and the noncategorical, needs-based system was introduced as early as 1987. However, the cornerstone of inclusive education policy came in 1996 with the introduction of Special Education 2000 (SE2000). SE2000 marked a paradigm shift, moving from an

individual deficit model to an ecological approach to addressing students' needs (Wills, 2006).

In inclusive education settings special needs (SN) students are placed in the same classroom as their peers and are provided with support and resources needed to succeed in their academic and social tasks. Inclusion of SN students in the classrooms brings multiple benefits to those students and society at large. Research shows that SN students who participated in mainstream classes have wider social networks and better social skills (Kvalsund & Bele, 2010), greater independence and employability as adults (Sun, 2007). Inclusion also improves societal acceptance of people with disabilities (Sun, 2007). At the same time a longitudinal study by Szumski et al. (2022) demonstrated that students without special needs achieve equally in inclusive and non-inclusive classrooms. Therefore, inclusive education is beneficial to both SN students and society without disadvantaging other students.

Placing students with different abilities in one classroom does not automatically lead to these students' equal participation in educational and social processes. In New Zealand, 8% of primary school-aged students and 15% of intermediate or secondary school-aged students have special needs (Mhuru, 2020). Thirty seven percent of students with additional needs attend special schools (Education Review Office, 2022). So, two thirds of SN students are included in mainstream schools, at least physically. However, SN students are 1.5 to 3 times more likely than their non-SN peers to face punitive measures, such as being stood down or suspended (Education Review Office, 2022). They also tend to change schools more frequently compared to non-SN students (Education Review Office, 2022). This highlights the urgent need for behaviour management strategies tailored to the needs of SN students. Additionally, while research shows that SN students in mainstream classes have more friends than those in special schools or classrooms, 40% of older SN students in New Zealand report difficulties in playing and forming friendships (Mhuru, 2020). This underscores the

importance of implementing additional strategies to support SN students in building social relationships.

Overall, students with special needs placed in general education classrooms are not necessarily included in classroom processes; such placement is hardly an inclusion and does not bring expected benefits to these students. Therefore, strategies must be implemented to support SN students so they could be properly included in classroom academic and social processes.

Response to Intervention

Response to Intervention (RTI) is an educational approach initially developed for special learning and inclusive settings with the focus on students with learning disabilities and their academic development. Like ABA, RTI focuses on the environment rather than the individual. It states that students' outcomes are influenced by classroom environment, curriculum and quality of instruction (Sugai & Horner, 2009). RTI initially focused on academic performance, mostly reading. However, as poor academic performance often goes hand in hand with behaviour difficulties, so RTI developed to address behaviours as well. The name "response to intervention" was borrowed from the medical field of toxicology where this phrase describes a person's reaction to emergency treatment (Sugai & Horner, 2009). Even though RTI borrows extensively from the medical field, it does not support the medical approach of excessively diagnosing people who do not cope under existing circumstances but focuses on how a certain student responds to a certain strategy and adjusting educational and behavioural strategies to the learner's needs to provide enough support for a student to thrive. RTI is both an educational strategy and a tool for identifying additional needs. Before any diagnosis is made, efforts must be made to provide adequate environment for a student's successful learning. Moving away from "one intervention – one school" approach RTI suggests to tailor education to students' needs (Sugai & Horner,

2009). RTI focuses on prevention of academic and behavioural difficulties and providing different levels of support.

Other important aspects of RTI are implementing only evidence-based practices, frequent progress monitoring by the assessments related to that student's curriculum, and data-based decision making. These are important considerations to ensure that students receive high-quality support. Assessing a student based on the student's curriculum ensures that knowledge and skills relevant to the individual learner are tested. For example, various research shows that IQ tests, frequently used as a diagnostic tool, consistently disadvantage students from minority groups as they prioritise knowledge and skills relevant to a different cultural and social group (Valencia & Suzuki, 2000). Progress monitoring and data-based decision making ensure that supports reflect current needs of the learner, and that achievement is recognised. Application of evidence-based support strategies ensures that no time is wasted on fads and the best strategies according to the current knowledge are implemented, reflecting the right of the individual to receive effective treatment (Sugai & Horner, 2009).

In order to support all students, a three-tiered system of early intervention and prevention was suggested. The first level of support (Tier 1) is universal support available for everyone and usually covers the needs of approximately 80% of students. Some students, approximately 20%, may not respond to Tier 1 support; then Tier 2 interventions are offered. Tier 2 supports are more intensive and often conducted in small groups, providing a higher teacher to student ratio and targeted teaching. Those students who do not respond to Tier 1 and Tier 2 supports, approximately 5%, are offered individual interventions. The major difference between a diagnostic discrepancy model and RTI approach is that a student receiving Tier 2 or Tier 3 supports is not labelled as a Tier 2 or Tier 3 student. This allows both students and teachers to avoid fixation on labels and choose supports according to

current needs of the student. For example, a student did not learn reading during whole class instruction and Tier 2 more structured small group instruction was provided for this student. As soon as this student learnt to read at the level of their classmates Tier 2 support was withdrawn. This example demonstrates that the student did not learn to read because of inadequate instruction and not because individual disability, which is indicated by the temporary nature of support intervention. If this student was not provided this additional support, the discrepancy between their and their peers' levels of academic achievement would increase which may lead to a learning disability diagnosis and possible behaviour issues. Therefore, the preventative approach of identifying the required support and providing it allowed this child to avoid a disability label and catch up with peers when the reason for poor academic success lay in inadequate instruction.

Schoolwide Positive Behaviour Interventions and Support

ABA concepts and knowledge and the RTI flexible approach to supporting students with additional needs were combined in the School Wide Positive Behaviour Interventions and Support (SWPBIS) framework. SWPBIS provides a comprehensive, multi-tiered behaviour support system for schools that teaches and encourages students' prosocial behaviour, ultimately improving their social and academic outcomes (Horner & Sugai, 2015; Lewis et al., 2010). Initially developed in the 1980s and 1990s in the United States, the Positive Behaviour Intervention and Support (PBIS) framework was a response to advocacy groups demanding more positive and preventive behavioural interventions for people with disabilities in community care and inclusive settings (Sugai & Horner, 2002). PBIS is grounded in behavioural theory and the principles and technology of applied behaviour analysis (ABA). PBIS uses the supportive actions of others to help individuals develop behaviours that lead to more productive and healthy lives. SWPBIS is an application of the PBIS framework specifically designed for educational settings, and it has achieved

widespread success with more than 25,000 schools in the USA implementing the framework (Center on PBIS, 2025).

Initially, four key principles of SWPBIS were identified (Sugai & Horner, 2002). The first principle is a strong focus on outcomes. The measurable outcomes for students and staff must be identified by the schools. The second principle of SWPBIS, is ensuring evidence-based practices are used for achieving these outcomes. The third principle states that decision making is informed by carefully collected data, highlighting its strong connection with ABA. The fourth principle of SWPBIS is its focus on systems (policies, committees, action planning etc.) that are needed to ensure the first three key principles can be achieved and a continuum of behavioural support is provided (Sugai & Horner, 2002). The fifth principle of equity was added more recently. The principle of equity emphasises the importance of local context and cultural considerations in order to enhance experiences of each student and staff member in school (Center on PBIS, 2025).

A distinctive feature of SWPBIS is that it considers the whole school as a unit of analysis. It is understood that the organisation does not "behave" but that the behaviours of each individual sum up to produce the collective outcomes. Therefore, addressing the behaviours of individuals systemically on various levels within the organisation leads to a certain group outcome (Sugai & Horner, 2002). Consistent with the RTI approach, SWPBIS offers a continuum of preventive behavioural support that is delivered in three levels, or tiers (Sugai & Horner, 2002). In school environments, these tiers, primary, secondary and tertiary, to some extent, correspond with systems of behaviour support. Tier 1 interventions aim to prevent problem behaviours from developing by implementing school wide strategies. Such strategies may include identifying and encouraging appropriate behaviours and organising opportunities to practice those behaviours, removing factors that encourage and maintain inappropriate behaviours, and maximising academic success. It is expected that

approximately 80% of students will respond to Tier 1 strategies. Tier 2 focuses on working with students who did not respond to Tier 1 interventions. Tier 2 interventions are more focused, delivered in small groups. For example, students with social difficulties can be provided with social skills training in a small group. The secondary tier of behavioural support addresses the needs of approximately 15% of students. The tertiary tier of behavioural support is aimed to support the rest 5% of students who failed to respond to Tier 2 interventions by providing individually developed and FBA-based support addressing complex and well-established problem behaviours (Sugai & Horner, 2002). Consistent with the RTI approach, students receiving a certain support are not labelled according to the level of support they received, emphasising the source of behaviour difficulties in the environment and not within the individual student, and the flexibility of tiered support.

SWPBIS was implemented in New Zealand in 2010. The Ministry of Education adopted Positive Behaviour for Learning (PB4L) policy signifying the shift towards positive behaviour management in the New Zealand education system. The application of PB4L in New Zealand was conducted in close collaboration with the developers of PBIS in the United States (New Zealand Ministry of Education, 2015). Initially, 87 schools participated in the programme. By 2013 more than 400 school adopted this framework. Although developed overseas, this framework showed positive results in New Zealand in terms of a significant decrease in stand down rates and increased NCEA level 1 achievement rates following two years of implementation of PB4L (Boyd & Felgate, 2015).

Efforts to adapt SWPBIS to the New Zealand context were made. For example, it was ensured that PB4L has “meaning for students, connects with their wider lives, and engages the support of their families, whānau and communities”. Whānau and communities were included as one of the PB4L systems and are regularly consulted about their values and implementation of new initiatives (New Zealand Ministry of Education, 2015). Teachers

within the PB4L framework are encouraged to strengthen their cultural self-awareness, knowledge and competence. Tools such as Tātaiako: Cultural Competences of Teachers of Māori Learners were developed.

Overall, SWPBIS provides a comprehensive framework for promoting appropriate behaviour and improving social and academic outcomes through a multi-tiered support system. Grounded in ABA and RTI principles, it emphasises proactive, evidence-based strategies, data-driven decision-making, and equity. Adapted to the New Zealand context as PB4L, this framework has shown promising results, including reduced disciplinary actions and improved academic achievements. By incorporating cultural considerations and fostering strong connections with whānau and communities, PB4L ensures its relevance and effectiveness in New Zealand schools, offering a flexible and inclusive approach to behaviour management.

Practices of the PBIS Framework

As indicated earlier, SWPBIS provides a framework for implementing ABA-based practices across various contexts in schools to create a continuum of behavioural support. Although practices and contexts can vary, there are several underlying principles that are consistent. These are: establishing clear expectations, explicitly teaching expected behaviours, reinforcing those behaviours, and discouraging inappropriate behaviours. One way to establish clear expectations is by developing school-wide values and providing concrete examples of expected behaviours and actions. In traditional approaches students may encounter different teachers communicating different values and expected behaviours, which can be confusing for them. When expectations are clear and consistent across contexts, they are easier to remember and follow for students.

Teaching expected behaviours is particularly important in social situations. Social skills are complex, and while some children easily pick up the rules and adapt, for many, this

does not occur naturally—such as for children with Autism Spectrum Disorder (ASD). Additionally, children from minority cultures and lower socioeconomic (SES) backgrounds may lack exposure to the social situations that most students encounter before starting school. Those who struggle to acquire social skills early on will further struggle with developing relationships and have limited opportunities to develop their social abilities. Therefore, in PBIS framework it is important that children are taught values, rules and appropriate behaviours explicitly.

Gresham et al. (2001) identify three stages in the social skills learning hierarchy: acquisition, performance, and fluency. While clear expectations help students understand social skills, further development and fluency require opportunities to practice these skills in natural social environments. Without regular practice, students may struggle to use social skills effortlessly and appropriately (Crewdson et al., 2024). To support this, a range of strategies and practices can be implemented in school settings to encourage and reinforce positive behaviours. One of the most common reinforcing practices is praise. Praise is a way to acknowledge and approve of a person's appropriate actions or correct responding (Brophy, 1981). Teachers can use general and behaviour specific praise (BSP). Although research shows that general praise is used more often, BSP shows better results in improving students' behaviour. BSP is a well-established educational practice with a robust evidence base supporting its effectiveness in promoting positive student behaviours (Ennis et al., 2020). BSP can be delivered immediately after the behaviour happened and, unlike general praise, specifies exactly which action was praised (Ennis et al., 2020). Therefore, BSP is more likely to function as a reinforcer as according to behavioural principles a consequence delivered immediately after behaviour occurred is more likely to increase the likelihood of behaviour occurring again than a consequence delivered after a delay. Praise is a very practical reinforcer as it does not require monetary expenses or special arrangement. It also does not

take much time and effort from the teacher making it a very practical and low-intensity intervention (Alberto & Troutman, 2013). Research shows that teachers who demonstrate high levels of praise rated themselves as more efficacious with regards to classroom behaviour management (Reinke et al., 2013). Additionally, when teachers use more praise, they use less reprimands (Reinke et al., 2007; Stormont et al., 2007).

However, despite multiple benefits, strong evidence base, and teacher development programmes teaching about the effectiveness of praise, research shows that levels of praise in classrooms are consistently lower than the recommended ratio of 3 or 4 praise statements to 1 reprimand. Rates of praise are the highest during the first year of school, then drop dramatically in the second year and continue to decrease further (Reddy et al., 2013). Shernoff et al. (2020) identified that barriers for implementing praise in classrooms include teachers' beliefs or training, loss of instructional time and challenges to using praise consistently. From a behavioural perspective, praise may not meet natural reinforcement whereas reprimands often do. For instance, if a student is screaming and the teacher reprimands them, resulting in the screaming stopping, the teacher experiences an immediate change in the environment, reinforcing their use of reprimands. Conversely, if a student is sitting quietly and the teacher praises them for their behaviour, no notable environmental change occurs for the teacher, leaving their behaviour unrewarded. This highlights the challenge of maintaining the beneficial rates of praise in the classroom and the need to explore other strategies to ensure students receive adequate reinforcement for their appropriate behaviours, even when teachers cannot provide it consistently.

Discouraging inappropriate behaviours is also part of the behaviour support continuum. In the PBIS framework, strategies such as pre-correction and planned ignoring are used to discourage inappropriate behaviours. Pre-correction is an antecedent strategy where an aspect of the environment is manipulated before the behaviour is about to happen.

For example, students are reminded about the rules in a high-risk situation before any rule-breaking occurred. Planned ignoring is an extinction strategy, where problem behaviour is ignored by the teacher. The combination of antecedent and consequence-based strategies allows to maximise the effectiveness of behaviour management and reduce negative side-effects.

Evidence-Based Classroom Programmes

Good Behaviour Game

The Good Behaviour Game (GBG) is a programme that uses interdependent group contingencies to discourage inappropriate behaviour. It was first introduced by Barrish et al. (1969) to reduce disruptive behaviours in elementary school classrooms. Later, it was applied as a preventative measure (Kellam et al., 1994). Foundational research by Kellam et al. (1994) demonstrated long-term positive effects on aggressive and disruptive behaviours through a large-scale epidemiological study.

In the Good Behaviour Game students are assigned to teams. Each team is given points for inappropriate behaviours of their members by the teacher and the team with the fewest points is rewarded (interdependent group contingency) (Barrish et al., 1969). Multiple teams can win if they meet the set criteria (for example five points or less). Over the time different adaptations of GBG were developed such as giving points for appropriate rather than inappropriate behaviours (Crouch et al., 1985), incorporating a merit system to enhance academic engagement (Schakel & Darveaux, 1984), adding a behavioural intervention (Wright & McCurdy, 2012), including a self-monitoring element (Babyak et al., 2000), not using teams (Harris & Sherman, 1973), independent and dependent group contingencies (Gresham & Gresham, 1982), and allowing individual students to earn points (Babyak et al., 2000).

Over the years research demonstrated that GBG is effective for a variety of issues, such as verbal and physical aggression, noncompliance, oppositional behaviours, hyperactivity, and out-of-seat behaviours (Bowman-Perrott et al., 2016). The GBG has also been associated with increases in prosocial behaviours like on-task behaviours, assignment completion, acceptance of authority, and improved concentration (Bowman-Perrott et al., 2016). Positive outcomes have been noted in both general education and special education settings from elementary to high school levels of education (Bowman-Perrott et al., 2016).

However, there are some drawbacks to the traditional GBG. First, while intending to encourage appropriate behaviours, GBG still focuses on inappropriate behaviours. With such focus the awareness of inappropriate behaviours in the classroom is heightened and perception about student behaviour may be unbalanced. Second, GBG is teacher led, therefore active awareness of problem behaviours has to be maintained by the teacher at all times taking teachers' time and energy and being potentially problematic if the teacher makes a mistake in judgement, similar to the punitive approach. Therefore, traditional GBG does not align with the principles of positive behaviour support.

The variations of GBG were developed to give points for appropriate behaviours instead of subtracting points for inappropriate ones (Crouch et al., 1985), and to incorporate an additional behavioural intervention (Ruiz-Olivares et al., 2010). In their meta-analysis, Bowman-Perrott et al. (2016) found that both original and modified formats of GBG were effective, offering flexibility of implementation. New, more positive variations of GBG fit the SWPBIS framework while being as effective as a traditional one in reducing problem behaviours.

Cooperative Learning

Cooperative learning (CL) is a widely recognised educational strategy based on students working in small groups towards a common academic goal. By emphasizing

collaboration rather than competition, CL fosters an environment in which students can grow both academically and personally, improving their self-esteem, communication skills, and respect for one another (Davidson, 1990; Madden & Slavin, 1983).

Key elements of CL are positive interdependence, individual accountability, heterogenous grouping, group goals and success (Gillies, 2016). Positive interdependence means that students perceive that their success is intertwined with the success of their group. Such interdependence creates an environment of collaboration encouraging students to develop mutual respect and cooperative effort (Johnson & Johnson, 1992). Individual accountability ensures that all members of a group contribute to the group's success, thereby promoting a sense of responsibility and engagement (Slavin, 1991). Heterogeneous grouping by the teacher ensures that groups include members of different abilities and in the event of a group competition all groups have similar chances of success. Working in such groups ensures the input from every student, encourages the exchange of different perspectives and fosters elaborate thinking (Davidson, 1990). Group goals are developed in such a way that they cannot be performed by students individually ensuring team effort.

Research consistently supports the effectiveness of cooperative learning in promoting student success. Studies indicate that students working in cooperative settings outperform those in traditional classroom settings, even in subjects like mathematics (Davidson, 1990). Further, CL has been found to be particularly advantageous for students with special needs, providing them with a supportive environment where they can contribute meaningfully to the group and develop social skills (Madden & Slavin, 1983; Slavin et al., 1994). High-achieving students also benefit from CL, as the collaborative setting allows them to challenge their thinking, refine their ideas, and develop a deeper understanding of the content (Johnson & Johnson, 1992). At the same time, researchers emphasise that CL works only when groups of students are rewarded for their actions (Slavin, 1991). Speaking in behavioural terms, CL is

effective only when group contingencies are in place. The carefully developed structure of CL ensures that desirable behaviours, such as collaborative behaviours, are reinforced by group contingencies.

There are three main ways to deliver rewards in groups: independent, dependent and interdependent group contingencies. When independent group contingencies are applied, the contingencies for a behaviour of an individual are delivered only to this individual but the contingencies are the same for all members of the group (Cooper et al., 2020). In dependent group contingencies, the reward is delivered to all group members contingent on the success of one (Cooper et al., 2020). When interdependent group contingencies are implemented, no individual member can access the reward alone. The whole group receives a reward only if the whole group reached the goal or fulfilled the task (Cooper et al., 2020). In this group contingency members of the group depend on each other. Interdependent group contingency is the one that is used in many school interventions and specifically in CL.

Despite the numerous advantages, implementing cooperative learning effectively requires significant efforts from teachers. Classroom activities must be carefully planned and organised. Teachers must structure activities to ensure that every student is actively engaged, and the goals are properly developed. Teachers also need to be able to manage group dynamics (Slavin et al., 1994). This includes the ability to facilitate productive interactions, guide group discussions, and provide appropriate feedback. Teachers also need to ensure that each student's contributions are valued and that all students are equally responsible for the group's success (Johnson & Johnson, 1992). Furthermore, successful implementation of CL requires an understanding of the diverse needs of students, as cultural and educational contexts may influence the effectiveness of cooperative learning (Sapon-Shevin, 1994).

Therefore, although cooperative learning is a great tool for improving student learning, collaboration skills and overall classroom environment, the level of teacher proficiency may present a barrier for implementation.

Positive Peer Reporting

Positive peer reporting (PPR) is a peer-mediated intervention designed to improve students' social relationships and increase positive behaviours (Morrison & Jones, 2007). The idea behind PPR is to provide an alternative to tattling and encourage students to "catch peers being good" (Haydon et al., 2023). During a specified time, usually at the end of the day, students publicly report the pro-social behaviours of one student who is randomly chosen as the star of the week (Grieger et al., 1976). In PPR, students learn to notice pro-social behaviours of their peers and give behaviour specific praise (although not immediate).

Peer reporting interventions have been successfully implemented in a wide range of educational settings from pre-school (Grieger et al., 1976; Smith et al., 2009) to high school (Lum et al., 2017) as well as in residential facilities and foster care settings (Ervin et al., 1996; Horn, 2004), establishing it as an intervention effective in a wide variety of settings and across a wide range of ages. In the SWPBIS framework peer reporting interventions are usually implemented in Tier 1 (Collins et al., 2018) though some authors suggest implementing PPR as a Tier 2 strategy in PBIS framework (Collins et al., 2018; Morrison & Jones, 2007).

One advantage of PPR as a peer-mediated intervention is that students can observe their peers' pro-social behaviours when teachers cannot. This ensures that more students' good behaviours are noticed. As mentioned earlier, research consistently shows that teachers use less than optimal praise (Reinke et al., 2013) and this intervention can help provide additional praise needed for the optimal balance. Therefore, PPR helps to increase students' and teachers' awareness of students' good behaviours, develop social skills and reinforce

prosocial behaviours. Additionally, as pro-social behaviours of the students can be noticed by their peers beyond the classroom, such as on a playground, it allows the desired behaviours to be generalised to other settings (Skinner et al., 2000)

Another advantage of PPR as a peer-mediated intervention is that by employing students this intervention places few demands on teachers. Most of the work of observing and reporting pro-social behaviours is conducted by students. The teachers' task is to organise the procedure. Additionally, unlike cooperative learning, PPR does not require special knowledge and skills. The main disadvantage of the original PPR intervention is that only one child at a time (the star of the week) directly benefits from the intervention and for a large class whole rotation may take a long time.

Tootling

Tootling is a variation of pro-social peer reporting (PPR) where students get the opportunity to be praised more often. The term Tootling combines "tattling" and the phrase "tooting one's own horn." Like PPR, Tootling encourages students to acknowledge and report their peers' pro-social actions. A traditional Tootling intervention includes three key components: peer reporting (Tootling), interdependent group contingencies, and public posting. These elements work together to create a supportive classroom environment that promotes and reinforces positive behaviour among students. Similar to PPR, Tootling employs students as the agents of change, saving teacher time and effort, and does not require the teacher to have specialised knowledge and skills.

Public posting is an element that is not present in CL and PPR. Public posting provides students feedback on their progress toward the goal and serves as a reminder that the programme is running prompting students to tootle. Using prompts and feedback to create behaviour change is a well-established practice in ABA (Cengher et al., 2018; Drabman & Lahey, 1974; Mangiapanello & Hemmes, 2015; Tekin-Iftar et al., 2019).

Initially, Tootling was developed by (Skinner et al., 2000) for use in general primary classrooms. The traditional Tootling procedure is implemented by teachers throughout the school day. At the beginning of the day, each student receives a reporting card that is placed on their desk. Students can fill out their cards anytime they observe a peer demonstrating appropriate behaviour. The reporting cards include spaces to record the name of the peer being reported, the name of the student submitting the report, and a description of the observed behaviour. Once a student has used their card, they can collect another one from the teacher's desk.

At the end of the day, the teacher reviews all submitted cards to ensure they meet the criteria and reads aloud five random cards to provide daily feedback to some students. The teacher also counts the total number of reports submitted that day and updates a poster chart showing the class's progress towards a collective goal. This goal is typically designed so that children can earn rewards every few days. The rewards, decided collaboratively by the teacher and the students, are usually inexpensive but highly meaningful to the children, such as extra recess time. Once the goal is achieved, the students receive their agreed-upon reward, and the progress chart resets to zero.

Various studies demonstrated the effectiveness of Tootling in improving student behaviour across various settings and ages. Research on Tootling is provided in greater detail.

Tootling Research

Initial studies by Skinner et al. (2000) and Cashwell et al. (2001) explored the effectiveness of the combination of interdependent group contingencies and public posting on increasing positive peer reports. The seminal study by Skinner et al. (2000) used an ABAB withdrawal design in a fourth-grade classroom. Although this study demonstrated some effect of the intervention on positive peer reports, unplanned punishment from the principal during the main intervention phase undermined the researchers’ efforts, resulting in weak

experimental control. A study by Cashwell et al. (2001) aimed to replicate the research by Skinner et al. (2000) in a second-grade classroom. It demonstrated strong experimental control and clearly illustrated the effect of reinforcing contingencies on pro-social peer reporting. Researchers also observed anecdotal evidence supporting the social and applied validity of the intervention, such as students cheering at the beginning of the second intervention phase. Additionally, the fact that the teacher continued the intervention for 2.5 weeks after the research was withdrawn indicated acceptability by the teacher and that Tootling not only increased positive peer reports but also had meaningful and practical impacts on classroom dynamics. As a result of these two studies, interdependent group contingencies and public posting became essential elements of the Tootling intervention.

Once the strategy was established, a major focus of Toolting research has been on evaluating the effect of Tootling intervention on classroom behaviour. Starting with Cihak et al. (2009), most studies evaluated its effect on disruptive behaviours (Chaffee et al., 2020; Dillon et al., 2019; Harry et al., 2023; Lambert et al., 2015; Lipscomb et al., 2018; Lum et al., 2019; Lum et al., 2017; McHugh et al., 2016; Powell, 2020). Lambert et al. (2015) were the first to evaluate appropriate classroom behaviours and other researchers followed (Chaffee et al., 2020; Derieux, 2019; Dillon et al., 2019; Harry et al., 2023; Lum et al., 2019; Lum et al., 2017; McHugh et al., 2016; Powell, 2020). Disruptive behaviour was the primary measure in most studies, setting the criteria for phase change, except for Harry et al. (2023) where academically engaged behaviour (AEB) was used as their primary measure and the basis for phase changes.

Tootling Across Settings

In twenty-five years of research, Tootling was implemented in various settings such as pre-school, all school grade levels, after school programmes, post-secondary education, general, and special classrooms. Miller (2017) implemented Tootling intervention in pre-

school settings. A multiple baseline across four classrooms research design was used in this study. Considering the age of participants the intervention was adapted. For example, instead of writing a tootle children would report vocally to the teacher and the teacher would add a token to the tootle tower, which was empty at the beginning of a new cycle. The colour of the token corresponded with the type of behaviour reported. The goal was gradually increased so it took longer each time to achieve. This was done to promote maintenance effects. However, after the first goal was achieved the rate of Tootling dropped and, in order to strengthen the behaviour the researchers decided to decrease the goal levels and set goals in consultation with the teachers. The results demonstrated the decrease in disruptive behaviours, increase in on-task behaviours and decrease in tattling. After the intervention was withdrawn and before the follow up the teachers had a choice whether they wanted to continue the intervention or stop it completely. Most teachers continued to implement some aspects of intervention but with much less fidelity. Follow up data showed some maintenance of reduced disruptive behaviour, increases in on-task behaviour, increases in tattling in three classrooms and decreases in tattling in the fourth classroom. Additionally, the overall teacher's stress, as measured by Teacher Stress Inventory (Fimian, 1988) reduced following intervention. The largest decrease in stress levels occurred in the 'Discipline and Motivation' subscale. The only subscale which showed increased score was the 'Time Management' subscale. Authors suggested that this could be due to the intensity of the Tootling intervention and difficulty with implementing all of the components every day. Overall, Tootling showed positive results in the preschool classroom.

While initial studies were focused on primary school children, later studies included middle school students (Chaffee et al., 2020), high school students (Lum et al., 2019; Lum et al., 2017), adults with disabilities in post-secondary education (Lipscomb et al., 2018), and children in after-school programmes for at-risk children (Kirkpatrick et al., 2019).

Some researchers suggested that public praise could be aversive for adolescents (Bear & Manning, 2008). Therefore, implementing Tootling in high school classrooms could be challenging. On the other hand, an intervention that can potentially restructure peer influence is even more needed in high schools. Lum et al. (2017) conducted a study in three general high school classrooms, evaluating class-wide disruptive (DB) and academically engaged behaviours (AEB), using an ABAB withdrawal design with a 10-second momentary time sampling recording procedure. Modifications were made to the Tootling procedure to make it more acceptable for high school students. For example, following classroom votes the intervention was renamed "Shout outs", “To-be-honests”, and “To-be-realz” for different classrooms. During the intervention phases, DB decreased, and AEBs increased. For two classrooms, the difference was very clear. The effect in the third classroom was not as pronounced but still meaningful according to statistical measures. This study demonstrated that positive peer reporting indeed reinforces pro-social behaviours, and that Tootling can be effective for high school students.

Chaffee et al. (2020) conducted a study with middle school classrooms and implemented a maintenance procedure. Two sixth-grade general education classrooms from a US middle school participated. The researchers employed a reversal design with a maintenance phase (ABABC) to measure class-wide disruptive and academically engaged behaviours. Similar to Lum et al. (2017), Chaffee et al. (2020) adapted some aspects of Tootling for secondary school students. These adjustments included presenting the intervention as a competition, voting on a name for the programme, and referring to peer reports as “positive comments” instead of “tootles”. To promote maintenance, the interdependent group contingency was removed during the final phase of the intervention.

The results indicated an increasing trend for academically engaged behaviours and a decreasing trend for disruptive behaviours in the first classroom. The levels for both

behaviours did not return to baseline numbers after the first implementation, and there was a notable overlap between phases. This outcome is similar to findings by Lum et al. (2017), suggesting that such patterns may be characteristic of secondary schools. The overlap between phases could indicate that secondary school students are more likely to adjust their behaviours even without a contrived reward. Chaffee et al. (2020) demonstrated that behaviour levels were maintained even when the reward was withdrawn, marking the first demonstration of a maintenance effect in Tootling research. However, these findings should be interpreted with caution, as the effect was also partially maintained during withdrawal phases. Further research is needed to investigate the lack of maintenance in primary school classrooms, where behavioural effects seem to be more sensitive to phase changes.

Lipscomb et al. (2018) implemented Tootling in a post-secondary classroom for adults with disabilities. It is the only study implementing Toolting with this age group. Participants were first-year students enrolled in a full-time Comprehensive Transition Programme (CTP) for adults with disabilities at a university. The author compared Tootling with ClassDojo to ClassDojo alone. ClassDojo is a computer and mobile application providing feedback about student behaviour to parents. This tool allows a teacher to report and track student behaviour as well as share the information with parents. In Tootling with ClassDojo, ClassDojo was used for peer reporting, counting and publicly displaying the reports. Using an alternating treatment design following the baseline phase, data on disruptive behaviours were collected for each of the seven students and classwide. The results suggested that both interventions were effective, but the ClassDojo condition resulted in slightly greater behaviour change. The authors speculated that this might be due to the ClassDojo condition involving teacher-reported and monitored behaviour, potentially causing students to behave better compared to the Tootling plus ClassDojo condition, where behaviours were reported by peers.

Nonetheless, this study demonstrated the effectiveness of the Tootling intervention and expanded the research to adults with disabilities in post-secondary education.

Some researchers also explored Tootling for special education. Thus, Ray (2019) implemented Tootling in a primary special education classroom. On-task behaviour, disruptive behaviour and pro-social behaviour (a positive social interaction with another student) were evaluated using a multiple baseline design across three classrooms. During this study Tootling occurred at a low rate. Two classrooms were able to achieve their goals (five tootles) once. One classroom never achieved the goal producing only four tootles. Despite low rates of reporting, on-task behaviour increased, and disruptive behaviour decreased in all three classrooms. One of the classrooms demonstrated increases in pro-social behaviour (a positive interaction). The author suggested that it could be due to the fact that students in this classroom played board games and card games during the majority of the study. This study demonstrated that students with social, emotional, or behavioural difficulties who study in special education classrooms may also benefit from the Tootling intervention.

Component Research in Tootling

Skinner et al. (2000) implemented a Tootling procedure as a combination of positive peer reporting using notecards, progress-tracking with a poster, feedback through the loud reading of five notecards, and an interdependent group contingency for reaching a goal. Consequent research generally retained this structure. However, due to the implementation of Tootling as a package it is impossible to attribute the success of the intervention to a single component. Goss (2019) indicated that there was criticism of rewards in some psychology literature arguing that it can reduce children's 'intrinsic motivation' and make their behaviour dependent of the external rewards (Bear et al., 2017; Kohn, 2006; Lillard, 2005; Montessori Jr, 1976). Additionally, considering the immediate effect that Tootling can have on peer relationships it could be expected that positive peer reporting in itself can motivate children

to continue practicing it. Therefore, Goss (2019) implemented Tootling in a lower elementary Montessori classroom without goals and rewards. In this study, children read and discussed stories about pro-social behaviours, before Tootling was implemented. This ensured that teacher's expectations were communicated to the students. The AB research design without withdrawal was chosen to ensure consistency for children. The results showed that the highest number of tootles appeared in the first week, likely due to a novelty effect. Then the number of reports reduced dramatically even if the teacher reminded the children to look for good behaviours and write tootles. No effect of the intervention on children's behaviours was demonstrated. There was a slight improvement in children's perception of their and their peers' behaviours. Although this study should be interpreted with caution due to its weak research design, it indicated that Tootling intervention implemented without rewards may be ineffective. Therefore, group contingency is an essential part of the package.

McHugh et al. (2016) examined whether an intervention with daily reinforcement would be more effective than the traditional approach, where it takes a few days for the class to reach the goal and receive reinforcement. Three primary classrooms (second and third grade) participated in this study, with one student in each classroom who exhibited more disruptive behaviours than others chosen as a target student. Therefore, data were collected for classrooms and individual target students. An ABAB withdrawal design with a multiple baseline element across two classrooms was utilised. Consistent with previous research, class-wide disruptive behaviour decreased, and academically engaged behaviour increased following the Tootling intervention. Data from the target students also indicated the desired effect of the Tootling intervention on their behaviour. Contrary to the hypothesis, the intervention with daily rewards was not superior to the traditional approach.

Derieux (2019) also aimed to evaluate the effectiveness of a single component of the Tootling package, specifically the contribution of report writing in Tootling process. The

study utilised an alternating treatment design with multiple baselines across three high school classrooms. The three alternating conditions were a traditional Tootling intervention, a comparison writing procedure, and a no-treatment control condition. Class-wide levels of disruptive and academically engaged behaviours were measured. In the comparison writing treatment, students wrote and returned notecards with two interesting things they learned that day and a goal for receiving a group reward was set by the teacher, similar to traditional Tootling. In the no-treatment control condition, students also submitted cards with their learnings of the day, but no reward was given. The results did not show significant differences in levels of AEB and DB between the conditions and the evidence for all the treatment conditions was mixed. This can suggest that the peer praise component of the Tootling intervention does not significantly contribute to the intervention's success. However, considering that the findings on Tootling effects on behaviour is not consistent with previous research, more research is needed to confirm Derieux (2019) findings.

In the original Tootling intervention, only five tootles are read aloud in the end of the day. This might be a small number of students who receive praise for their pro-social behaviours. Harry et al. (2023) explored whether enhancing the Tootling intervention by rewarding students reported for pro-social behaviours through more detailed public posting would increase its effectiveness. Conducted in four high-school classrooms in rural US, this study focused on academically engaged behaviour as the primary dependent variable. Using a multiple baseline design with A/B/B+C phases, where B represented Tootling intervention with reduced feedback (no reading tootles aloud) and B+C represented Tootling with modified public posting, the study found increased academically engaged behaviour and decreased disruptive and passive off-task behaviours during both B and B+C phases. Tootling with increased public posting demonstrated larger effects and more stable data. However, authors reported that the difference in the effects between the variations was marginal.

Despite the hypothesis on modified public posting not being supported, this study contributes to the evidence on the effectiveness of Tootling intervention for high school classrooms.

Public posting and interdependent group contingencies have been essential elements of the Tootling intervention since its inception. However, the steps of counting tootles and updating the chart complicate the procedure. This complexity requires more time and effort from teachers, which can lead to low levels of treatment integrity, as demonstrated in previous research (Lambert et al., 2015; Lum et al., 2017; McHugh et al., 2016). Additionally, the effort required may result in teachers' unwillingness to continue the intervention despite its positive outcomes. To address this issue, Lum et al. (2019) suggested using a randomised independent group contingency instead of the traditional interdependent group contingency. In this modified approach, a total of five rewards were distributed daily. Three tootles were randomly picked from the tootle box, and the three students whose behaviour was reported received the rewards. Additionally, two names were picked from a second box containing the names of students who submitted tootles. This intervention demonstrated an increase in class-wide academically engaged behaviour and a decrease in disruptive behaviour. Passive off-task behaviour, a new measure introduced in this study, remained low in all phases and did not change with the intervention. Overall, such modification was successful in replicating the effects of traditional Tootling procedure while simplifying the process for the teacher. One drawback of this approach, though, is that it necessitates the use of tangible reinforcements adding to the cost of the intervention.

Kirkpatrick et al. (2019) investigated the use of the modified reward criteria on individual student behaviours. The author tracked antisocial and disruptive interactions of four teacher-nominated students in an after-school programme for at-risk children. An ABAB withdrawal design was used. Unlike in previous studies, the reward criteria were tied to the number of different students being tootled about, rather than the total number of tootles. This

alteration aimed to motivate students to tootle about a variety of peers and, therefore, encourage all students to exhibit pro-social behaviours. The results of this study demonstrated the effectiveness of this Tootling variant in decreasing antisocial and disruptive interactions among the target students which is consistent with other Tootling literature.

Tootling and Social Skills Training

While some researchers modified components of Tootling, others investigated combining Tootling with other interventions. Three studies (Crewdson et al., 2024; Wright et al., 2022; Wright et al., 2021) have integrated Tootling with social skills training (SST), using Tootling to reinforce newly taught social skills. Previous research has demonstrated successful skill acquisition following this SST. However, there is limited evidence showing that children consistently use and maintain these skills in their social groups over time. Considering that Tootling was created as a reinforcement environment for incidental social skills learning, it could function similarly for recently taught social skills. Wright et al. (2021) explored this concept by using Tootling to reinforce the skill of giving compliments among first-grade students attending an after-school programme for at-risk students. The Tootling intervention was modified to better suit the study's purpose. Specifically, students reported only the compliment-giving behaviour of their peers, and the intervention was named the "Catching Compliments Game" (CCG). The study employed an ABAB withdrawal design. Children were divided into groups to participate in activities, and both group-wide and class-wide compliment-giving behaviour was measured. The results demonstrated a clear effect of CCG on increasing compliment-giving behaviour both class-wide and within groups. This research suggests that Tootling can effectively reinforce recently taught social skills, helping children to apply and maintain them in their social interactions.

Building upon the work of Wright et al. (2021), Crewdson et al. (2024) conducted a study, focusing on two recently taught social skills within an after-school first-grade

classroom for at-risk students. Students were instructed to report instances of compliments and encouragement as part of the intervention. Evaluation encompassed both class-wide and group-specific assessments of these behaviours within an ABAB withdrawal design. Results indicated a substantial increase in the targeted social skills during the Tootling intervention phase, with a corresponding decline upon withdrawal of the intervention. These findings align with and broaden the conclusions drawn by Wright et al. (2021) regarding the efficacy of Tootling in reinforcing recently acquired social skills.

Extending the research on strengthening Social Skills Training with Tootling intervention, Wright et al. (2022) conducted a study introducing three pro-social behaviours sequentially within the Tootling framework. Employing an interdependent group contingency with randomly selected criteria, the study aimed to encourage students to continue demonstrating previously taught behaviours as new ones were introduced. The targeted behaviours included compliments, encouragement, and "thank you" statements. Results indicated a considerable increase and maintenance of all target behaviours throughout the phases, demonstrating the effectiveness of the procedure for teaching children new skills while maintaining previously acquired ones. However, "thank you" statements exhibited some variance between phases, notably increasing when compliment giving was introduced. This suggests a natural response from students and emphasizes the importance of carefully considering the order in which behaviours are introduced for optimal experimental control.

Interaction of Tootling with Classroom Activities

During their research on SST with Tootling (Wright et al., 2022), the authors noticed that some activities evoked higher levels of social interactions than others, therefore providing more opportunities to engage in target behaviours. In order to test this hypothesis Crewdson et al. (2023) conducted a study exploring the interaction between the types of activity and the levels of target behaviours. Two types of activity (Lego® blocks and Yeti in

my Spaghetti® game) and two types of social skills (compliment and encouragement) were chosen for the study. Fifth-grade students attending an after-school programme for at-risk children participated in this study. An alternating treatment design was chosen for this study with alternating Lego®, Yeti® and Bingo conditions where Bingo served as a control. The results showed that Yeti was superior to both control and Lego® in terms of the amount of both compliments and encouragements produced whereas the Lego® condition only produced compliments and not encouragements. Children did not demonstrate any target behaviours in the control condition. Although such results could be expected, the authors highlighted the need to consider activities children are involved in when Tootling is in place. The types of activities children are involved in can significantly affect the number of opportunities children have to engage in desired behaviours.

Tootling and Technology

In line with recent interest in using technology in education, Tootling researchers also explored the opportunity to incorporate technology in Tootling interventions. Dillon et al. (2019) implemented the Tootling procedure via ClassDojo in three fifth-grade classrooms. Tootling sessions were twenty minutes long, during which students were encouraged to take mental notes of their peers' good behaviours. At the end of the session, students could submit their tootles via a ClassDojo application on a classroom computer. The researchers found this procedure to be very time-efficient; even the largest class with the fewest computers was able to submit all their tootles within three minutes. The study employed an ABAB withdrawal design to measure disruptive and academically engaged behaviours, showing a considerable effect of the intervention. Researchers noted that behaviours quickly returned to baseline levels following the withdrawal of the intervention, which resulted in lower effectiveness scores from the teachers and negatively impacted the overall social validity score. Therefore, alongside simplifying the procedure, researchers should develop maintenance strategies to

ensure lasting behaviour change and secure teacher buy-in. Simplifying the procedure did not come without drawbacks. With less control from the teacher, students were caught Tootling on absent students or on themselves.

Tootling and Teacher Behaviour

Most Tootling research to date is focused of student behaviour. To evaluate the effect of Tootling on teacher behaviour, Powell (2020) implemented a traditional Tootling intervention in three primary New Zealand classrooms using a multiple baseline design with a follow up. In addition to on-task behaviour, the author evaluated teachers' corrective and praise statements. Corrective statements were used to get insight into students' disruptions during class time. The intervention caused the expected effect on student behaviour, increasing on task and decreasing disruptions in one group as measured by teacher's corrective statements. No or even a slight reverse effect was observed on teachers' praise statements, which reduced as intervention progressed. Although both teachers refused to participate in a maintenance procedure for this study, one teacher implemented natural fading, conducting Tootling for most, but not all, days following the end of the main part of the study and before the follow-up. As expected, levels of student behaviours in this class were consistent with the intervention phase. Levels of student behaviours in another class, where Tootling was not continued by the teacher, returned to baseline levels. Powell (2020) also evaluated the behaviours of target students which were chosen in each class. All three target students were chosen in consultation with the teachers as they exhibited higher levels of disruptive behaviours than their peers. Results for those students were consistent with previous findings and classwide data of this study. Anecdotally, one student reported to his Resource Teacher that Tootling helped him with improving his behaviour. Overall, Powell (2020) was the first study to implement Tootling in New Zealand and the first one to evaluate teacher behaviour as well as student behaviour.

Building upon Powell (2020), Choundira (2022) replicated the study in inclusive primary and middle New Zealand classrooms evaluating the effect of the intervention on academically engaged and disruptive behaviours of individual students at risk for emotional and behavioural disorders (EBD) and comparison groups of randomly selected students. Modified reinforcement criteria tied to the number of students tootled (Kirkpatrick et al., 2019) were used in this study to encourage students to tootle on different peers. Teacher praise statements were also evaluated during this study. The results demonstrated an increase in academically engaged behaviour in two of three students and increase in teacher praise in one of two teachers. This study was conducted as a series of two AB designs due to restrictions of Covid-19 pandemic. Although, such design does not allow for replication of the effect and any causal conclusions it adds some evidence to the effectiveness of Tootling intervention especially for inclusive classrooms in New Zealand.

Enright (2023) also evaluated teacher behaviour in a study conducted in a primary New Zealand classroom. The study was conducted using ABABCD withdrawal design with fading and follow up. The results on student behaviour demonstrated increases in instruction-following and on-task behaviours in two of three students and their peers. Consistent with Powell (2020), the results on teacher praise were mixed. The results on teacher stress as measured via the Teacher Stress Inventory (TSI) were also mixed. However, Enright (2023) argued that teacher stress subscales should be interpreted separately. Three subscales which showed the increase of stress level are ‘Work-Related Stressors’, ‘Professional Distress’ and ‘Professional Investment’. These subscales usually reflect stress associated with the larger school environment such as large workload, overload with administrative tasks and paperwork. This stress could be explained by the changes in teachers’ work related to COVID-19. For example, during the time of research, the teacher had to cater for both in class and online students. Two subscales that showed decrease in stress levels were

'Discipline and Motivation' and 'Time Management'. These subscales are more related to student behaviours and teachers' feeling in control of classroom environment. Therefore, more research is needed on teacher praise, and although the results regarding teacher stress are mixed, they could indicate promising opportunities. Further research is needed during more stable times to reach definitive conclusions about the effects of the Tootling intervention on teacher behaviour.

Gaps in Tootling Research

One consistent gap in Tootling research is the lack of implementation of maintenance and generalisation procedures. This gap is frequently highlighted in studies but remains largely unaddressed. Teachers often reject the idea of implementing a maintenance procedure, as it significantly increases the length of a study and the researchers' presence in the classroom. Consequently, only a few studies have reported the implementation of maintenance procedures, and one study noted that a teacher unintentionally carried out such a procedure. Including maintenance procedures as part of a Tootling intervention is crucial, as this can help sustain behavioural gains with less effort from teachers. If Tootling, as an antecedent intervention, typically influences participants' behaviour while it is actively implemented, then the behavioural improvements would disappear once the intervention ends. Research supports this observation, showing that behaviour levels usually return to baseline during withdrawal phases or at the conclusion of the intervention. In some cases, this decline occurs gradually over several sessions. Studies that included maintenance procedures have demonstrated that the positive effects of Tootling can be sustained when a deliberate strategy is in place. Furthermore, evidence suggests that the impact of maintenance procedures may vary across different age groups. Given the potential benefits of well-designed maintenance procedures for enhancing Tootling interventions, further research in

this area is essential, particularly across diverse demographic groups and with long-term follow ups.

Although there are several studies that explored modification of key components of Toolting, there is limited comparison between modified and traditional Tootling. For example, in an attempt to evaluate a more detailed public posting Harry et al. (2023) compared their modified procedure not with a traditional one, where daily feedback is given to students by reading five reports aloud, but to another modified procedure where this daily feedback is withdrawn. Therefore, the modified procedure with reduced feedback was compared with another modified procedure with increased feedback. Other examples are implementing independent group contingencies and using ClassDojo, both of which demonstrated being effective interventions however were not compared with a traditional Tootling procedure. Comparison of Tootling variations between each other can be useful to assist decision-making for teachers or other decision-makers when choosing the variant of Toolting for their classroom or school.

Collins et al. (2020) in their meta-analysis of peer-reporting interventions indicated that measures of school climate are rarely included in peer-reporting studies and none of the Tootling studies included in the meta-analysis reported any. Considering that Tootling is believed to improve peer interactions it would be natural that it improves classroom social environment. Therefore, measures of student perceptions of each other and their teacher, belongingness to the school and social support could be evaluated (Collins et al., 2020). Goss (2019) evaluated children's perceptions of each other and reported that they improved after Tootling was implemented. However, this study had an AB research design and Tootling procedure, implemented without rewards, was ineffective. It would be especially beneficial to understand the perceptions of special needs and at-risk children on school climate when

Tootling is implemented. Therefore, research in this area is needed across various school demographics.

Purpose of This Study and Research Questions

Over the past two decades, research on Tootling has evolved, becoming more complex and nuanced. Originally a simple peer-reporting strategy, Tootling has expanded into an intervention with various adaptations for diverse environments and needs. Tootling is now implemented and studied across different student populations and is evaluated not only for its impact on peer reporting but also on broader classroom behaviours for both students and teachers. Therefore, it is important that Tootling research is reviewed with the aim of finding out if it can be considered an evidence-based strategy. However, few studies have synthesised evidence specifically regarding Tootling's effectiveness as a stand-alone strategy. In meta-analyses, Tootling was combined with other positive peer-reporting (PPR) interventions, limiting the depth of analysis on its independent effects. The first review to cover Tootling studies was conducted by Murphy and Zlomke (2014). They reviewed twenty-four studies to identify common components of peer reporting interventions, with no specific analysis of PPR's effects on students' behaviour. Only two of those studies involved Tootling.

A more recent meta-analysis by Collins et al. (2020) aimed to systematically analyse peer-reporting interventions, with thirteen of the twenty-one studies having Tootling as their independent variable. The authors examined the characteristics of PPR interventions, their overall effects, and the variables that contribute to their effectiveness. They concluded that PPR strategies positively impact student behaviour, reducing disruptive behaviour and improving social engagement and academic involvement. The study found no major differences in effectiveness between traditional PPR and Tootling. Despite these positive findings, the analysis gave limited attention to the effects of Tootling (and other PPRs) across varied populations. Grade levels were assessed broadly, with categories of preschool/primary

and secondary/postsecondary, but no separate analysis was performed for students with special needs, possibly due to insufficient data at the time. As Tootling gains further attention from educators and researchers, several new studies have emerged between 2020 and 2024, offering additional insights into the effectiveness and applicability of Tootling across different educational settings and student populations.

It was suggested that Tootling can be used as a Tier 1 or Tier 2 intervention within the SWPBIS framework (Lum et al., 2017; McHugh et al., 2016). However, Tier 2 interventions are conducted in small, focused groups (Center on PBIS, 2025); therefore to receive a Tier 2 intervention target students need to be removed from their classrooms and placed in a different group for the time of the intervention. Since Tootling supports target students without removing them from the classroom or singling them out in any other way, it may be more beneficial if used as a Tier 1 strategy. Implementing Tootling at this level can promote the inclusion of students with special needs in mainstream education, potentially increasing the proportion of students whose needs are met within Tier 1 supports. Therefore, the purpose of the current study was to investigate if Tootling supports the inclusion of students with special needs in general education classrooms. To serve this purpose, a systematic review of the extant Tootling literature was conducted, and the research questions were formulated as follows:

1. Is Tootling effective in improving classroom behaviours in general education classrooms that include students with special needs?
2. Is Tootling an evidence-based strategy for enhancing classroom behaviours in general education settings that include students with special needs?
3. Is Tootling effective in improving classroom behaviours specifically for students with special needs?

4. Is Tootling an evidence-based strategy for improving classroom behaviours for students with special needs?

Method

Database Search

A database search was conducted using the PsycINFO, Scopus, Education Search Complete and Google Scholar databases. The keyword used in all searches was "Tootling". The query parameters limited results to English-language documents published between 2009 and 2024. In PsycINFO, additional restrictions were applied to include only quantitative studies. The Scopus search was limited to the Psychology subject area and the document type was specified as ‘article’. No additional restrictions were applied to Education Search Complete. The search in Google Scholar was performed using advanced search settings. The option ‘Include citations’ was unselected and the restriction of the key word being present in the article titles was added. The PsycINFO search returned 24 records, the Scopus search returned 19 records and the Education Search Complete returned 15 records. Google Scholar yielded 54 records, resulting in an initial total of 112 records.

Literature Selection

The records discovered by the database search were evaluated against the following inclusion criteria:

1. Studies needed to be experimental and quantitative; reviews, posters, meta-analyses, books, and other non-experimental works were not included.
2. Theses, dissertations, and other non-peer-reviewed works that were later published as peer-reviewed articles were excluded.
3. Studies were required to take place in general education primary or secondary classrooms.
4. For studies reporting classroom-level data, participants needed to be general education classrooms with students who had diagnosed special needs or received individual support. For studies reporting student-level data, participants needed to

be individual students with diagnosed special needs or those receiving individual support.

5. Tootling had to be the primary independent variable.
6. Dependent variables included at least one of the following target behaviours: disruptive behaviours (DB), appropriate behaviours (AB), academic engagement (AE), academically engaged behaviours (AEB), on-task behaviours (OTB), or passive off-task behaviours (POT).

The selection process was conducted in steps. The first step involved removing duplicates, with 32 duplicates automatically removed by the citation software EndNote and an additional 20 titles removed manually, resulting in 60 unique records at the end of the identification process. Next, abstracts and full texts of the articles were screened, and records that did not present experimental studies were removed. Sixteen books, reviews, posters, meta-analyses, and other non-experimental records were excluded. Further, seven theses and dissertations were removed due to the availability of peer-reviewed articles based on the same studies, leaving 36 studies. Finally, the remaining 36 studies were reviewed to determine if they met inclusion criteria 3 to 6. After applying these criteria, 17 studies were excluded for not meeting the final inclusion criteria, leaving 19 studies for further analysis. From these, only classrooms and individuals meeting Criterion 4 were included in the analysis. However, in studies using a multiple baseline design, all classrooms within the design were evaluated for quality, as they were included in the systematic visual analysis.

Quality Assessment

Finally, studies that reached the last stage of selection were evaluated for quality using the What Works Clearinghouse (WWC) quality standards for single-case design (SCD) studies. The WWC, an initiative developed by the U.S. Department of Education, aims to analyse educational research and provide reliable, meaningful guidance to education

providers (Institute of Education Science, 2025). Its guidelines offer a structured framework for assessing research quality and have been widely applied in various fields, including behavioural research that utilises SCD. The pilot version (Version 1.0) of the WWC standards for SCD studies was first issued as a separate document in 2010 (Kratochwill et al., 2010). It was later updated and integrated into a broader handbook covering multiple research approaches while maintaining its core principles. However, Version 4.1 introduced substantial changes to the SCD standards, which remained consistent in Version 5.0. (Institute of Education Studies, 2020, 2022). The primary aim of Versions 4.1 and 5.0 is to align the evaluation of SCD studies with conventional large-sample experimental designs (Kratochwill et al., 2021). As a result, some of the quality criteria introduced in these versions are not directly relevant to SCD research (Kratochwill et al., 2021). Additionally, the newer standards rely on the calculation of the design-compatible effect size (D-CES), a metric developed to equate SCD studies with large-sample research. The “5-3-20” threshold for combining studies was abolished. Kratochwill et al. (2021) argued that this approach is more suitable for meta-analysis than for reviews. Therefore, this review follows Version 4.0 of the *WWC Procedures and Standards Handbook* (Institute of Education Studies, 2017).

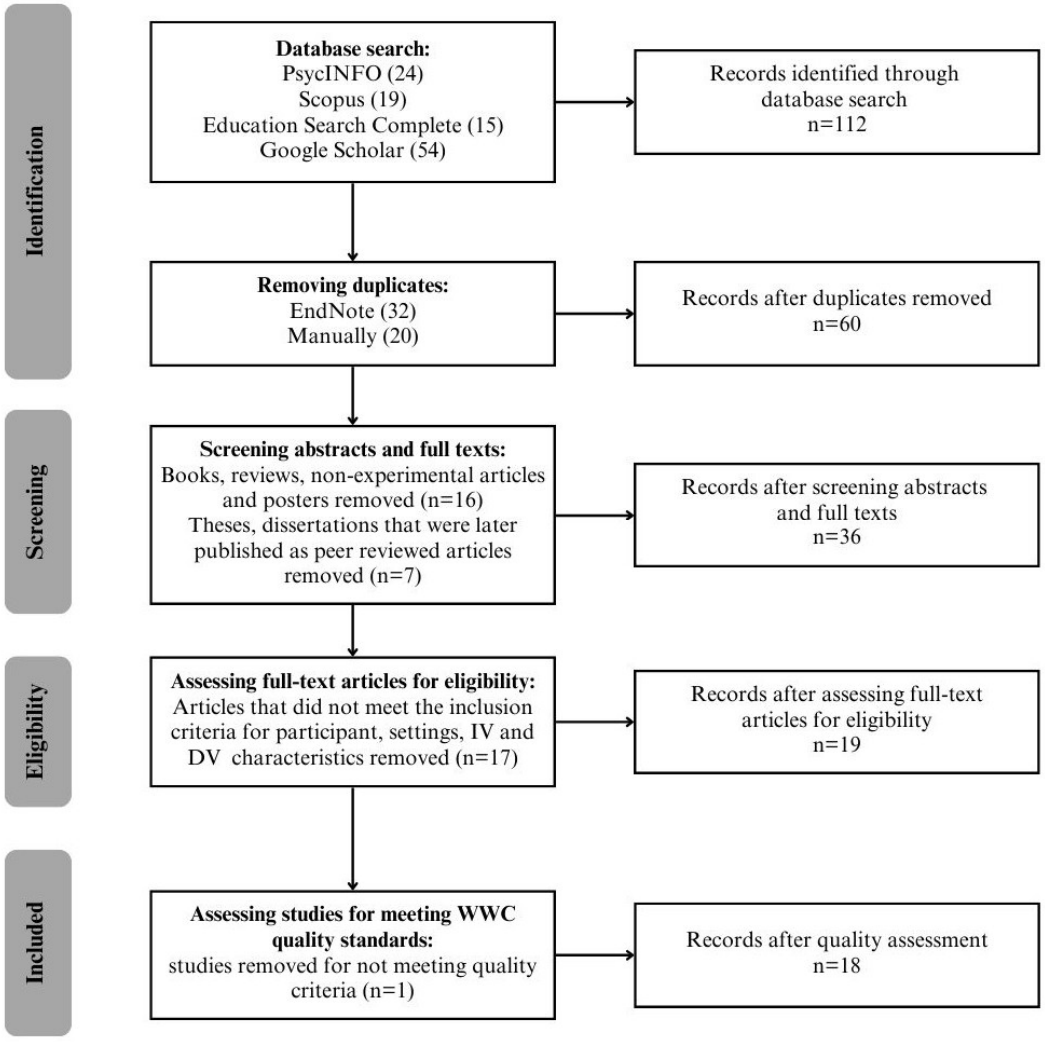
According to WWC standards (Institute of Education Studies, 2017), studies were classified into three categories: "Meets Standards," "Meets Standards with Reservations," and "Does Not Meet Standards". In order to meet the standards, studies had to meet the following criteria (Institute of Education Studies, 2017): (a) the independent variable was systematically manipulated, with the researcher deciding when and how conditions should change; (b) inter-observer agreement (IOA) was reported for every phase, covering at least 20% of data points in each condition, with a minimum of 80% agreement threshold; (c) the study attempted to demonstrate the effects of the independent variable at least three times at different points in time; (d) the study included at least three data points per phase for withdrawal and multiple

baseline designs and at least four data points per condition for alternating treatment designs. For a withdrawal research design to contain three attempts of demonstration of the effect, a minimum of four phases is required. The simplest example of a withdrawal design that meets WWC criteria is an ABAB research design. Multiple baseline designs require a minimum of six phases. The most common MBD that meets WWC criteria is a design with three baselines with two phases in each. Studies with five data points per phase meet the standards fully, while studies with three data points per phase meet the standards with reservations (Institute of Education Studies, 2017). These quality criteria are usually applied to the baseline and intervention phases, excluding maintenance phases and follow-ups. After quality assessment was completed, 18 studies met standards with or without reservations while one study were excluded. Additionally, data from four studies were only partially included in the review. In Chaffee et al. (2020), data from one of the two classrooms did not meet the standards, so only the data from the other classroom was included in the analysis. Similarly, in Enright (2023), data for one of the three students did not meet the standards, and only the remaining two students were included in the review. In Powell (2020), while class-wide observations met the standards, individual student data did not and was therefore excluded. In Tran (2024), IOA for inappropriate noise in the "total tootles" condition did not meet the standards, leading to the exclusion of this condition from further analysis. Consequently, only the portions of data meeting the standards from these studies were included in this review, rather than excluding the studies in their entirety.

The literature search and selection process is illustrated in the PRISMA diagram (Liberati et al., 2009) in Figure 1.

Figure 1.

PRISMA diagram illustrating literature search and selection process.



Quality assessment results, based on WWC research standards, are provided in Table 1 and Appendices B and C.

Table 1.

Evaluation of studies according to WWC quality standards.

#	Study	Level of analysis	Participants	Research design	Study meets WWC standards
1	Cihak et al. (2009)	Classroom	n/a	Withdrawal	with reservations
2	Lambert (2014)	Classroom	Classroom A	Withdrawal	with reservations
3	Lambert et al. (2015)	Classroom	Classroom B	Withdrawal	yes
4	McHugh et al. (2016)	Classroom	Classroom B	Withdrawal	yes
		Classroom	Classroom C	Withdrawal	with reservations
5	Lum et al. (2017)	Classroom	Classroom A	Withdrawal	yes
		Classroom	Classroom B	Withdrawal	yes
		Classroom	Classroom C	Withdrawal	yes
6	Derieux (2019)	Classroom	Classroom A	MB/AT	yes
		Classroom	Classroom B		
		Classroom	Classroom C		
7	Dillon et al. (2019)	Classroom	Classroom A	Withdrawal	yes
		Classroom	Classroom C	Withdrawal	yes
8	Lum et al. (2019)	Classroom	Classroom A	Withdrawal	yes
		Classroom	Classroom B	Withdrawal	yes
		Classroom	Classroom C	Withdrawal	with reservations
9	Wright (2019)	Classroom	Classroom A	Withdrawal	yes
		Classroom	Classroom B	Withdrawal	yes
		Classroom	Classroom C	Withdrawal	yes
10	Chaffee et al. (2020)	Classroom	Classroom A	Withdrawal	yes
		Classroom	Classroom B	Withdrawal	no
11	Powell (2020)	Classroom	Group 1	MB	yes
		Classroom	Group 2		
		Classroom	Group 3		
		Student	Nikau	Withdrawal	no
12	Choundira (2022)	Classroom	Classroom A	Withdrawal	no
13	Enright (2023)	Classroom	Classroom	Withdrawal	yes
		Student	Alea (TS3)	Withdrawal	yes
14	Harry et al. (2023)	Classroom	Classroom A	MB	yes
		Classroom	Classroom B		
		Classroom	Classroom C		
		Classroom	Classroom D		
15	Morris (2024)	Classroom	n/a	Withdrawal	with reservations
16	Tran (2024)	Classroom	n/a	AT	yes, except for inappropriate noise in the total tootles condition

Continued on the next page

Table 1.

Evaluation of studies according to WWC quality standards (cont.).

#	Study	Level of analysis	Participants	Research design	Study meets WWC standards
17	Whitefield et al. (2025)	Classroom	Classroom A Classroom B Classroom C	MB/AT	with reservations
18	Salinas (2021)	Student	Student A-1	Withdrawal	yes
		Student	Student A-2	Withdrawal	yes
		Student	Student B-1	Withdrawal	yes
		Student	Student B-2	Withdrawal	yes
		Student	Student C-1	Withdrawal	yes
		Student	Student C-2	Withdrawal	yes
19	Thoele and Sayeski (2024)	Student	Elian	AT	yes
		Student	Asher	AT	yes
		Student	Darnell	AT	yes
		Student	Jonah	AT	yes
		Student	Xavier	AT	yes

Data Extraction

The included studies were summarised based on the characteristics of the participants, variations of Tootling intervention, research details, and their findings. The characteristics of the participants include different levels such as country where the study was conducted, school characteristics, classroom and individual details, such as presence of special needs. Variations of Tootling interventions include how often and for how long it was implemented, any limitations on reporting, variations of public posting, use of technology, etc. Research details include research designs, observation methods, recording procedure and any additional measures such as social validity. For analysis, studies were categorised based on whether class-wide or individual student data were collected.

Research findings were evaluated in two ways. The first approach focused on the summarising the data on presence and size of the effect as reported in each study. The presence of the effect was documented for each study, and effect sizes were extracted from studies that

reported them. The aim of the review was to analyse the effect of Tootling across classrooms and students as well as across studies. This required separating multiple baseline design (MBD) studies and presenting their effect sizes individually. It is worth noting that all MBD studies in this review were conducted across participants (classrooms, students). Since classrooms from MBD studies include the first phase change in their effect size, taking the first phase change effect size value across all classrooms would provide the most consistency.

In addition to summarising research findings, a systematic visual analysis (SVA) was conducted as a complimentary evaluation in this review. WWC standards (Institute of Education Studies, 2017; Kratochwill et al., 2010) suggested conducting visual analysis as a second step of the review process after the assessment of quality. Although the full WWC review is beyond the scope of this work it was decided to conduct a SVA in order to obtain an additional perspective on the reported findings in order to come to more robust conclusions.

Visual analysis is traditionally the primary method for determining whether a functional relationship exists between an intervention and observed behaviour changes in single-case research (Kratochwill et al., 2013). In this process, graphed data are examined within and across experimental phases and compared to the data pattern that would be expected if no intervention had been implemented. The key components of visual analysis include the evaluation of level, trend, and variability of the data. Level refers to the position of the data on the y-axis (Barton et al., 2018). Trend describes the direction of data movement within a phase, which can be increasing, decreasing, or flat (Barton et al., 2018). Variability represents the spread of data points around the trendline (Barton et al., 2018). Changes in any of these characteristics may indicate an intervention effect. Replicating such changes across different points in time strengthens confidence in the functional relationship between the treatment and behaviour.

Visual analysis of data is used in behaviour analysis from its inception starting from B.F. Skinner's laboratory experiments. As the field developed further, more and more research has been conducted in clinical rather than laboratory settings with less control over the environment. It led to restrictions on the duration of the experimental phases and the data being 'messier' (Morley et al., 2017). Therefore, there are less opportunities for researchers to demonstrate clear effects. Some researchers have raised concerns about the low inter-rater agreement in visual analysis (Brossart et al., 2006; DeProspero & Cohen, 1979; Wolfe et al., 2016). The absence of formal step-by-step instructions allows for subjective influence, as factors such as an analyst's training, experience, and attentiveness may affect their conclusions (Fisch, 1998). To address this limitation, Wolfe et al. (2019) developed a systematic visual analysis protocol for withdrawal and multiple-baseline designs. This protocol uses a series of dichotomous questions to guide data interpretation, ultimately producing an experimental control score ranging from zero to five. The higher the score the stronger the evidence of a functional relationship between the independent and dependent variables. Interventions scoring below three are considered ineffective. This scoring system enables comparison and summarisation of study results without effect size calculations.

In this review, SVA scores were determined for all studies using withdrawal and multiple-baseline designs and reported in tables. These scores were also graphed alongside effect size descriptors extracted from the original studies to compare intervention effects. Since the SVA protocol evaluates multiple-baseline studies as a single unit, average comparison effect sizes were used. This involved averaging effect sizes across all phase changes in withdrawal designs and across all baselines in multiple-baseline studies. When studies did not report these averages, they were calculated by the researcher. This analysis, therefore, provides SVA scores for all withdrawal and multiple-baseline studies, along with a visual comparison of these scores against the effect sizes reported in the reviewed papers.

Interobserver Agreement

Inter-observer agreement (IOA) for research articles identification was collected for 30% of the articles at each stage of the literature selection by a second observer. The second observer was a second-year student in the Master of Applied Psychology (Behaviour Analysis) programme. The observer was provided with the instructions to conduct each step of the data selection process. IOA for the database search was not collected, as this step is time-sensitive and may yield different results depending on when it is conducted, even if all steps are followed correctly (e.g., if a new article is published). Instead, the observer was provided with a library of results from the data search conducted by the researcher and was instructed to follow all other steps of the data selection process. IOA was collected for 30% of the records at each step. Therefore, the observer was given randomly selected lists containing 18, 13, 11, and 6 records for Screening out duplicates, Screening out non-experimental articles, excluding thesis and dissertations that were published as peer-reviewed articles, and applying final eligibility criteria. After completing these steps, the observer was asked to identify the provided records in their resulting lists (See Appendices A, B, C, D for the forms used). IOA for study selection was calculated using the total agreement procedure (Johnston et al., 2020): smaller total ÷larger total ×100. The average IOA across all steps achieved between the researcher and the observer was 100%.

IOA was also measured for the quality assessment of the studies. An observer was provided with a list of classroom- and student-level studies comprising 30% of the total studies and was asked to evaluate them according to WWC quality standards (Institute of Education Studies, 2017). The form used for this IOA, along with the second observer's conclusions, can be found in Appendix E. IOA for quality assessment was calculated using the exact agreement formula (Johnston et al., 2020): (number of agreements ÷ total number of records) × 100%. The agreement for quality assessment was 100%.

Interobserver agreement was conducted for systematic visual analysis, covering at least 30% of all calculations for each combination of a target behaviour and a participant. This included eight scores for DB, eight scores for AB and two scores for POT in studies with classroom-level data. In studies with student-level data, IOA included two scores for DB and two scores for AB. IOA for each score was calculated using the modified total agreement formula (Johnston et al., 2020): (smaller value ÷ larger value) × 100. The average agreement was then determined by calculating the mean of the individual IOA values. The IOA for visual analysis in classroom-level studies was 90.86%, while for student-level studies, it was 95.00%. The forms used for the IOA for systematic visual analysis are presented in Appendices F and G.

Results

Data search, selection and quality assessment resulted in inclusion of 18 studies in this review. Selected studies presented data at two levels: classroom and student. Consequently, the results are reported in two streams based on these data levels. Most studies focused on a single data type; however, Enright (2023) provided both classroom and student data and was included in both analyses. Therefore, sixteen studies are included in the classroom-level stream, and three studies are included in the student-level stream.

The numerical identifiers were ascribed to each study in order to simplify the process of identifying studies across this section (see Table 2).

A variety of effect size measures were reported in the reviewed studies, including NAP, PND, Tau-U, and baseline-corrected Tau. Effect size descriptors such as "large," "moderate," and "small" were applied to improve consistency of interpretation across different effect size measures. Because in some studies descriptors were not provided and one study used an outdated interpretation for their effect size (Lum et al., 2017), interpretation strategies specific to the effect size measures in each study were used to assign descriptors. Thus, interpretation ranges for Tau-U (Vannest & Ninci, 2015), baseline-corrected Tau (Vannest & Ninci, 2015), PND (Scruggs et al., 1987), and NAP (Parker & Vannest, 2009) are presented in Table 3.

Studies with Classroom Level Data

Demographics of Classroom Level Studies

In total, 16 studies (studies 1 to 16) involving 30 participating classrooms with classroom-level data were analysed in this review. On average, classrooms in the studies had 21.60 students, with 3.87 of them being students with special needs. Authors of studies 11 and 12 did not specify the number of students receiving individual support in their classrooms. Therefore, three classrooms from those studies were not included in the calculations of the average number of students in the classroom.

Table 2.

The identifiers and full references of the reviewed studies.

Identifier	Study
1	Cihak, D. F., Kirk, E. R., & Boon, R. T. (2009). Effects of Classwide Positive Peer "Tootling" to Reduce the Disruptive Classroom Behaviors of Elementary Students with and without Disabilities. *Journal of Behavioral Education, 18*(4), 267-278.
2	Lambert, A. M. (2014). *Evaluating the Use of Tootling for Improving Upper Elementary/Middle School Students' Disruptive and Appropriate Behavior* [Doctoral Dissertation, The University of Southern Mississippi].
3	Lambert, A. M., Tingstrom, D. H., Sterling, H. E., Dufrene, B. A., & Lynne, S. (2015). Effects of Tootling on Classwide Disruptive and Appropriate Behavior of Upper-Elementary Students. *Behavior Modification, 39*(3), 413-430.
4	McHugh, M. B., Tingstrom, D. H., Radley, K. C., Barry, C. T., & Walker, K. M. (2016). Effects of Tootling on Classwide and Individual Disruptive and Academically Engaged Behavior of Lower-Elementary Students. *Behavioral Interventions, 31*(4), 332-354.
5	Lum, J. D. K., Tingstrom, D. H., Dufrene, B. A., Radley, K. C., & Lynne, S. (2017). Effects of Tootling on Classwide Disruptive and Academically Engaged Behavior of General-Education High School Students. *Psychology in the Schools, 54*(4), 370-384.
6	Derieux, J. (2019). *Is the Written Component of Tootling Effective? A Comparison of the Group Contingency to a Comparison Writing Procedure* [Master's Thesis, The University of Southern Mississippi].
7	Dillon, M. B. M., Radley, K. C., Tingstrom, D. H., Dart, E. H., & Barry, C. T. (2019). The Effects of Tootling via ClassDojo on Student Behavior in Elementary Classrooms. *School Psychology Review, 48*(1), 18-30.
8	Lum, J. D. K., Radley, K. C., Tingstrom, D. H., Dufrene, B. A., Olmi, D. J., & Wright, S. J. (2019). Tootling With a Randomized Independent Group Contingency to Improve High School Classwide Behavior. *Journal of Positive Behavior Interventions, 21*(2), 93-105.
9	Wright, S. J. (2019). *Peer-Mediated Tootling with a Standardized Form and a Mystery Motivator in High School Classrooms* [Doctoral Dissertation, The University of Southern Mississippi].

Continued on the next page

Table 2.

The identifiers and full references of the reviewed studies (cont.).

Identifier	Study
10	Chaffee, R. K., Briesch, A. M., Volpe, R. J., Johnson, A. H., & Dudley, L. (2020). Effects of a Class-Wide Positive Peer Reporting Intervention on Middle School Student Behavior. *Behavioral Disorders, 45*(4), 224-237.
11	Powell, C. (2020). *Use your kind words please: Effects of tootling on student and teacher behaviour in New Zealand primary-school classrooms* [Master's Thesis, The University of Waikato].
12	Enright, C. J. (2023). *An investigation into the effects of tootling for instruction-following and on-task student behaviour in a New Zealand primary school* [Master's Thesis, The University of Waikato].
13	Harry, S. W., Tingstrom, D. H., Dufrene, B. A., Dart, E. H., Radley, K. C., Lum, J. D. K., & Bates-Brantley, K. E. (2023). The Effects of Tootling Combined with Public Posting in High School Classrooms. *Journal of Behavioral Education, 32*(3), 565-584.
14	Morris, J. K. (2024). *The Effect of Tootling on Academic Engagement and Prosocial Behavior* [Ed.S, University of Nebraska at Omaha].
15	Tran, K. H. (2024). *A Comparison of Reward Criteria in Tootling* [Master's Thesis, California State University].
16	Whitefield, B. L., Harry, S. W., McIntosh, D. E., McIntosh, C. E., Chitiyo, A., English, A. M., & Poore, B. R. (2025). Elementary School Tootling: Implications on Notecard Responding and Teacher Observed Accuracy. Journal of applied school psychology, 41(1), 44-71.
17	Salinas, E. A. (2021). *Evaluating the Effects of Tootling on Disruptive and Academically Engaged Behaviors of Students with Emotional and Behavioral Disorders* [Doctoral Dissertation, The University of Texas at San Antonio].
18	Thoele, J. M., & Sayeski, K. L. (2024). The Effects of Student- and Teacher-Led Tootling on Student Disruptive Behavior. *Frontiers in Education, 9*.

Among the 30 classrooms reviewed, 27 were located in the US, and three classrooms were located in New Zealand (studies 11, 12). Fourteen classrooms were elementary/primary (studies 1-4, 7, 11, 12, 14, 16); one was middle (study 10); and 15 were high school level

(studies 5, 6, 8, 9, 13, 15), covering grades 1 to 12. These represented students aged approximately 5 years (a New Zealand primary classroom in study 11) to 18 years (U.S. high school classrooms in studies 5 and 8). Since most studies specified grade level rather than age, student ages were inferred based on each country's grade-level standards. The author of study 12 specified age but not the grade for their classroom, so the grade was inferred

Table 3.

Interpretation ranges for different effect sizes.

Effect Size Measure	Range	Interpretation
Tau-U, baseline-corrected Tau	> 0.80	Very large effect
	0.60 – 0.80	Large effect
	0.20 – 0.60	Moderate effect
	≤ 0.20	Small effect
PND	≥ 90%	Strong/large effect
	70 – 89%	Moderate effect
	50 – 69%	Questionable effect
	< 50%	No effect
NAP	0.93 – 1.00	Large effect
	0.66 – 0.92	Moderate effect
	0.00 – 0.65	Weak effect

according to New Zealand standards.

Five classrooms (from studies 2,3,4 and 12) were part of schools using the School-Wide Positive Behavioural Interventions and Supports (SWPBIS) framework, while 25 classrooms operated in non-PBIS schools. Classrooms included students with various special needs:

- Specific Learning Disability (SLD): studies 1, 2, 3, 5, 6, 7, 8, 9, 13, 16.
- Other Health Impairment (OHI/OHD): studies 4, 5, 6, 8, 9, 13, 16.
- Autism Spectrum Disorder (ASD)/Autism (AU): studies 5, 6, 7, 12, 13.
- Attention Deficit Hyperactivity Disorder (ADHD): studies 1, 7, 10.

- Emotional Behaviour Disorder (EBD)/ Emotional Disability (EMD): studies 6, 11, 16.
- Language/Speech Impairments: studies 13, 16.
- Visual Impairment: study 9.

Other supports for students reported in the studies include "pull-in" supports for reading instruction and in-resource learning support, as well as the development and implementation of individualized education plans (IEPs) for students (studies 14 and 15). The summary of demographic characteristics of the participants across classroom levels studies is presented in Table 4.

Table 4.

Demographic characteristics of classrooms from classroom level studies.

Category	Details	Studies, % of Total		Classrooms, % of Total	
Total Studies	16 studies				
Total Classrooms	30 classrooms				
Class Size Averages	21.60 students total, 3.87 with special needs				
Age Range	Ages 5 (NZ primary classroom) to 18 (U.S. high school)				
Grade Levels	Grades 1–12				
	Elementary/primary	9	56.3%	14	46.7%
	Middle	1	6.3%	1	3.3%
	High	6	37.5%	15	50.0%
Special Needs/ Supports	SLD	10	62.5%	20	66.7%
	OHI/OHD	7	43.8%	15	50.0%
	ASD/AU	5	31.3%	7	23.3%
	ADHD	3	18.8%	4	13.3%
	EBD, EMD	3	18.8%	5	16.7%
	Language/Speech Impairments	2	12.5%	3	10.0%
	Visual Impairment	1	6.3%	1	3.3%
	Pull-in services for reading comprehension	1	6.3%	1	3.3%
	In-resource learning support	1	6.3%	1	3.3%
	IEPs	1	6.3%	1	3.3%
Using SWPBIS Framework	Yes	4	25.0%	5	16.7%
	No	12	75.0%	25	83.3%
Locations of Studies	USA	14	87.5%	27	90.0%
	New Zealand	2	12.5%	3	10.0%

Research Characteristics of Classroom Level Studies

Of the 16 studies reviewed in the classroom level stream, 9 were peer-reviewed articles (studies 1, 3, 4, 5, 7, 8, 10, 13, 16), while the remaining 7 were non-peer-reviewed works (studies 2, 6, 9, 11, 12, 14, 15), including dissertations, theses, and one field project. Nine of the studies were conducted at the University of Southern Mississippi (studies 2-9, 13). Study 16, although conducted at a different institution, had one overlapping author with the University of Southern Mississippi team. Other contributing institutions included the University of Tennessee, the University of Georgia, Live Oak Behavioral Psychology, the Central Louisiana Resource Center on Developmental Disabilities, Washington State University, the Kennedy Krieger Institute, the University of Utah, Northeastern University, the University of California, the University of Waikato, Ball State University, the University of South Florida, Mississippi State University, the University of Nebraska, California State University, and the Vancouver School Board. Overall, the reviewed studies were carried out by six independent research teams without overlapping authorship across various locations in the United States and New Zealand.

All studies included in this review used Tootling as the primary independent variable. Twelve studies implemented a single variant (studies 1-5, 7-12, 14), three studies (studies 6, 13, 16) used two variants, one employed three variants (study 15). In study 6, traditional Tootling was compared to an alternative writing procedure and Tootling without rewards. While various student and teacher behaviours were measured, this review focuses on behaviours relevant to the research questions: disruptive behaviours (DB), appropriate behaviours (AB), academic engagement (AE), or academically engaged behaviours (AEB), on-task behaviours (OTB), passive off-task behaviours (POT), as defined by each study's authors. For the purpose of this analysis, AB, AE, AEB and OTB were grouped together under the umbrella term appropriate behaviours (AB) and the abbreviation AB will be used

further for all the above-mentioned behaviours. Other behaviours measured in the reviewed studies that were irrelevant to the research questions of this review, such as notecard responding, were not included in the analysis. DB was measured in 15 studies (all studies except study 12), AB in 14 studies (all studies except 1 and 15), POT in 3 (studies 6, 8, 13).

In terms of methodology, a withdrawal design was used in 11 studies (studies 1-5, 7 - 10, 12, 14), with fading procedures being incorporated in two (studies 10 and 12), and follow-ups included in four studies (studies 3, 5, 11, 12). Four studies (studies 6, 11, 13, 16) employed a multiple baseline (MB) design. Although studies 3 and 4 reported implementing a MBD, only two baselines were presented, which is not enough for a quality MB design according to WWC standards (Institute of Education Studies, 2017). Each tier in these studies had ABAB designs and were treated as separate ABAB designs instead of the tiers of the MBD. Three studies (studies 6, 15, 16) utilised an alternating treatments (AT) design. In studies 6 and 16 the AT design was part of an MB design with three baselines. In study 16, three Tootling variants were alternated, but the "total tootle" variant did not meet quality criteria and was excluded from the summary of the findings. As some studies employed hybrid research designs, the sum of the designs implemented is larger than the total number of studies.

Thirteen studies (studies 2 - 13, 16) conducted initial screenings of participants before the enrolment withholding participation for the classrooms and students with low levels of disruptive behaviours (lower than 20-30% depending on a study) and high levels of appropriate, academic or on-task behaviours (higher than 70-80% depending on a study) to prevent floor and ceiling effects.

Momentary time sampling was the predominant recording method, utilised in 12 studies (studies 2, 3, 5 - 13, 16). Other methods include interval recording (studies 4, 14), event recording (studies 1, 14, 15), and duration recording (study 15). Observation sessions

generally lasted around 20 minutes, though some extended up to 42 minutes, except for one study (study 1), in which observations were conducted by the teacher throughout the entire school day. Group observation methods varied: nine studies employed an individual-fixed method for group observation (studies 5 - 11, 13, 16), three observed all students simultaneously (studies 1, 14, 15), two separated students into rotating groups (studies 2 and 3), one used a rotation method without further details (study 4), and one selected two random students per session (study 12). Instead of observing the whole class Powell (2020) (study 11) observed seven students that were identified by the teacher as representative of the class.

Treatment integrity was measured in all 16 studies. Potential issues with treatment integrity were reported in study 13 as their teacher was completing integrity checklists only when the researcher was present (77% of the sessions). Although the IOA for these sessions was 100% it is unclear if the teacher regularly held the same standards of Tootling implementation. Thirteen studies included effect sizes in their reports (all studies except 1, 2, and 15). Effect sizes were calculated using various measures, with four studies applying NAP (studies 2, 4, 5, 10), 8 using Tau-U (studies 5, 8, 9, 10, 11, 12, 13, 16), two using baseline-corrected Tau (studies 6 and 7), and one employing PND (study 14). Notably, Lum (2017) (study 5) and Chaffee (2020) (study 10) calculated both NAP and Tau-U for effect sizes. As Tau-U was the most used effect size across the studies it was the one reported in findings for these studies for better consistency with other studies.

The intent for this analysis was to aggregate the data on Tootling effectiveness across classrooms despite their participation in withdrawal or MB designs. Therefore, the optimal consistency would be achieved if the effect sizes for the first phase change across all classrooms was reported. While some authors reported effect size values for two phase changes and a weighted average in their ABAB designs (e.g. studies 5, 8, 9, 10), others only reported an average value (e.g. studies 4 and 14). Therefore, due to the inability to extract

first-phase-change effect size values from averages, it was decided to use the average effect size values across all withdrawal design studies. Additionally, the weighted averages for Tootling without public posting and Tootling with public posting were presented in Study 12 for each target behaviour. These effect sizes were included in the analysis as they were, even though these values did not represent first-phase-change effect sizes or a single Tootling variant. Social validity among teachers was measured in 15 studies (all studies except study 14) and among students in 11 studies (studies 2, 4, 6, 8, 9 - 13, 16). The summary of the studies' research characteristics is presented in Table 5.

Tootling Procedure across Classroom Level Studies

There were many variations of the Tootling procedure across the studies. Only 4 studies (studies 1, 11, 13, 14) maintained whole-day implementation of the traditional Tootling, while others applied the intervention during blocks lasting 20 minutes to 2 hours (studies 2-10, 15, 16), or for half a school day (study 12). Two studies allowed students to submit tootles at any point during the day (studies 1 and 11); in four studies, tootles were returned during transitions (studies 6, 9, 13, 16); in five studies, they were returned anytime within a designated intervention block or lesson (studies 2-5, 8); in three studies, tootles were submitted at the end of the block or lesson (studies 7, 14, 15); and in one study, they were returned during lunch break (study 12). One study did not report timing for Tootling submissions (study 10).

Eight studies did not set a limit on the number or timing of reports (studies 1, 4, 5, 8, 10, 11, 14, 15). Tran (2024) (study 15), although initially setting a limit of one tootle per student, later removed it, hence it is considered a "no-limit" study. Submission limits of 1 to 3 tootles per student per session were imposed in 7 studies (studies 2, 3, 6, 9, 12, 13, 16). Authors of study 7, facilitated by Class Dojo technology, limited submission time to 5 seconds per student.

Table 5.

Research characteristics of classroom level studies.

Category	Details	Studies, % of Total		Classrooms, % of Total	
Publication Type	Peer-reviewed articles	9	56.3%	18	60.0%
	Non peer-reviewed works	7	43.8%	12	40.0%
Research Institutions	University of Southern Mississippi	9	56.3%	20	66.7%
	Other institutions	7	43.8%	10	33.3%
Research Teams	6 independent research teams				
Research Design	Withdrawal	11	68.8%	18	60.0%
	MB	4	25.0%	11	36.7%
	AT	3	18.8%	7	23.3%
Tootling Variants	Single variant	12	75.0%	20	66.7%
	Two variants	3	18.8%	9	30.0%
	Three variants	1	6.3%	1	3.3%
Dependent Variables	DB	15	93.8%	29	96.7%
	AB	14	87.5%	28	93.3%
	POT	3	18.8%	8	26.7%
Fading	Yes	2	12.5%	2	6.7%
	No	14	87.5%	28	93.3%
Follow-up	Yes	4	25.0%	7	23.3%
	No	12	75.0%	23	76.7%
Participant Screening	Yes	13	81.3%	26	86.7%
	No	3	18.8%	4	13.3%
Observation Method	Individual-fixed group observation	9	56.3%	22	73.3%
	Simultaneous observation of all students	3	18.8%	3	10.0%
	Rotating groups	2	12.5%	2	6.7%
	Random student selection (2 per session)	1	6.3%	1	3.3%
Recording Method	MTS	12	75.0%	25	83.3%
	Interval recording	2	12.5%	3	10.0%
	Event recording	3	18.8%	3	10.0%
	Duration recording	1	6.3%	1	3.3%
Observation Session Length	20 minutes - 42 minutes	15	93.8%	29	96.7%
	Full-day observations	1	6.3%	1	3.3%
Treatment Integrity Measured	Yes	16	100.0%	30	100.0%
Effect Size Calculation	NAP	4	25.0%	7	23.3%
	Tau-U	8	50.0%	18	60.0%
	Baseline corrected Tau	2	12.5%	5	16.7%
	PND	1	6.3%	1	3.3%
Social Validity	Teachers	15	93.8%	29	96.7%
	Students	11	68.8%	21	70.0%

Only two studies used technology in the intervention, both employing the Class Dojo app on school computers (studies 7 and 14). Standardized Tootling cards that allowed students to tick behaviours they wanted to report were developed in three studies (studies 7, 9, 12). Two variants of reporting cards, traditional and standardised, were utilised in Study 16 as two alternating treatments.

In terms of goal-setting, some researchers specified the number of days for children to reach a goal and receive a reward. Six studies set a goal reachable in 1-2 days (studies 1, 4, 6, 8, 15, 16), while 6 studies designed goals to be reached in 3 days or more and up to two weeks (studies 2, 3, 5, 7, 12, 13). The design of Tootling procedure in study 12 allowed students to reach their goal in one day but to receive a reward it was required to reach the goal three times. Randomisation was introduced in study 9, allowing daily goal achievement, with a lottery to determine if the class received a reward. Authors of the studies 10, 11, 14 did not specify their goal setting strategy. In all studies except study 6 students were consulted when choosing the rewards for reaching their goals. However, it was not reported in study 16 if the students were consulted. Students in six studies were given a chance to vote for the intervention name (studies 5, 8, 9, 10, 13, and 15).

Thirteen studies (studies 1-7, 9-11, 13, 14, 16) used a total tootle count as the goal, while 2 linked the goal to the number of students submitting tootles (studies 8 and 15) and 3 based it on the number of students reported by their peers (studies 8, 12, 15). Lum (2019) and Tran (2024) (studies 8 and 15), incorporated different goal strategies in different phases of their study. Goals were known by students in all studies except study 12, where it was kept secret.

In terms of public acknowledgement of children's reporting, 11 studies included reading 5 tootles aloud at the end of the session (studies 1-5, 8 – 12, 16), while in one study (study 7) all tootles were displayed on a screen, and in one study (study 13) all tootles were

posted publicly on a wall in one of their two conditions while no acknowledgement was implemented in the other. Public acknowledgement of any form was not implemented in three studies (studies 6, 14, 15). Progress towards goals was displayed publicly in 13 studies (studies 1-7, 10 – 14, 16). An interdependent group contingency was used in 15 studies (studies 1-7, 9-16) and only one used an independent group contingency (study 8). The main interventionist was typically the teacher (13 studies: 1-5, 7, 8, 10-14, 16), though 2 studies involved teacher-researcher collaboration (studies 6, 15), and 1 study employed students as interventionists (study 9). Variations of Tootling procedure across reviewed studies is presented in Table 6.

Findings across Classroom Level Studies

The following summary reflects findings as reported by the authors of the evaluated studies. DB was evaluated in 29 classrooms (studies 1-11, 13 - 16) and Tootling was reported to be effective in improving DB in 25 classrooms (studies 1-11, 13 - 16). Among the 26 classrooms for which effect sizes were reported, 17 classrooms demonstrated high, large and very large effects (studies 3-5, 7, 8, 13, 14, 16), 6 classrooms demonstrated moderate effects (studies 6, 9, 10, 16), while only 3 classrooms showed small effects (studies 6 and 11). AB was evaluated in 28 classrooms (studies 2-14, 16) and Tootling was reported to be effective in 27 of those classrooms (studies 2-16). The only classroom that did not demonstrate the effect of Tootling on AB belonged to study 6. Among 27 studies which reported effect size, 22 classrooms demonstrated high, large and very large effects (studies 3-14, 16), 4 classrooms demonstrated moderate effects (studies 5, 6, 9, 13), and only 1 classroom showed a small effect (study 6). The effect of Tootling on POT was explored in 3 studies with 8 classrooms all of which reported effect sizes (studies 6, 8, 13). The effect was demonstrated in 6

Table 6.

Variations of Tootling procedure across classroom level studies.

Category	Details	Studies, % of Total		Classrooms, % of Total	
Implementation Duration	Whole-day traditional Tootling	4	25.0%	7	23.3%
	Blocks of 30 minutes to 2 hours or half-day	12	75.0%	23	76.7%
Timing of Tootling Submissions	At any point during the day	2	12.5%	3	10.0%
	During transitions	4	25.0%	12	40.0%
	Anytime within designated blocks/lessons	5	31.3%	9	30.0%
	At the end of a block/lesson	3	18.8%	4	13.3%
	During lunch break	1	6.3%	1	3.3%
	Timing not reported	1	6.3%	1	3.3%
Submission Limits	No limits on number/timing of reports	8	50.0%	13	43.3%
	Limited to 1–3 tootles per session	7	43.8%	12	40.0%
	Submission time limited to 5 seconds	1	6.3%	2	6.7%
Use of Technology	Class Dojo app on school computers	2	12.5%	3	10.0%
Tootling Card Types	Traditional cards	11	68.8%	18	60.0%
	Standardized cards with tick options	3	18.8%	6	20.0%
	Both standardized and free-text options	1	6.3%	3	10.0%
Voting for Intervention Name	Yes	6	37.5%	13	43.3%
	No	10	62.5%	17	56.7%
Student Input in Choosing Rewards	Yes	15	93.8%	27	90.0%
	No	1	6.3%	3	10.0%
Goal-Setting Strategies	Goals achievable in 1–2 days	6	37.5%	12	40.0%
	Goals set for 3+ days (up to 2 weeks)	6	37.5%	11	36.7%
	Randomized daily goals with lottery	1	6.3%	3	10.0%
	Not reported	3	18.8%	4	13.3%
Goal Criteria	Total tootle count	13	81.3%	26	86.7%
	Number of students submitting tootles	2	12.5%	3	10.0%
	Number of students reported by peers	3	18.8%	4	13.3%
Goal Transparency	Goal known to students	15	93.8%	29	96.7%
	Goal kept secret	1	6.3%	1	3.3%
Public posting of progress	Yes	13	81.3%	24	80.0%
	No	3	18.8%	6	20.0%

Continued on the next page

Table 6.

Variations of Tootling procedure across classroom level studies (cont.).

Category	Details	Studies, % of Total		Classrooms, % of Total	
Public Acknowledgment of Reports	Reading 5 tootles aloud at session end	11	68.8%	20	66.7%
	Displaying all tootles on a screen	1	6.3%	2	6.7%
	Posting all tootles publicly on a wall	1	6.3%	3	10.0%
	No public acknowledgement	3	18.8%	5	16.7%
Type of Group Contingency	Interdependent group contingency	15	93.8%	28	93.3%
	Independent group contingency	1	6.3%	2	6.7%
Tootling Implementers	Teacher	13	81.3%	23	76.7%
	Teacher-researcher collaboration	2	12.5%	4	13.3%
	Teacher and students	1	6.3%	3	10.0%

classrooms with 2 classrooms showing large and very large effect sizes (study 13) and 4 classrooms showing moderate effect sizes (studies 6, 8, 13). Small effect size was reported for two classrooms in studies 6 and 8. Effectiveness of Tootling procedure across classroom-level studies is summarised in Table 7.

Systematic Visual Analysis of Classroom Level Studies

Following the protocol outlined by Wolfe et al. (2019), data from classrooms with ABAB withdrawal designs were evaluated individually producing a score for each classroom, while data from classrooms in multiple baseline designs (MBD) were processed as one lot producing a single score for that study. As a result, the term "classrooms" refers to both individual classrooms in ABAB designs and aggregated classrooms in MBD designs. A positive effect of Tootling on DB was identified in 19 (studies 1-5, 7-11, 16) out of 22 classrooms that measured DB (studies 1-11, 13, 14, 16); and a positive effect of Tootling on AB was identified in 20 (studies 2-5, 7-12, 14, 16) out of 22 classrooms (studies 2-14, 16)

Table 7.

Effectiveness of Tootling across target behaviours in classroom level studies.

Target behaviour	Evaluated	Effective		Effect Size Reported	High/Large/Very Large Effect, %		Moderate Effect, %		Small/No Effect, %	
	Class	Class	%	Class	Class	%	Class	%	Class	%
DB	29	25	86.2%	26	17	58.6%	6	20.7%	3	10.3%
AB	28	27	96.4%	27	22	78.6%	4	14.3%	1	3.6%
POT	8	6	75.0%	8	2	25.0%	4	50.0%	2	25.0%

that measured AB. The effect of Tootling on POT was evaluated in 5 classrooms but was not found to be effective in any of those classrooms (studies 6, 8, 13). The results of SVA for classroom-level studies are summarised in Table 8.

Further, the systematic visual analysis scores were compared with effects size descriptors according to effect sizes reported in the studies for the DB and AB as target behaviours with the largest amount of data. Because the studies with MBD designs produced

Table 8.

Effectiveness of Tootling based on systematic visual analysis score in classroom level studies.

Target Behaviour	Total Classrooms Evaluated	Effect Demonstrated		Effect Not Demonstrated	
		Classrooms	%	Classrooms	%
DB	22	19	86.4%	3	13.6%
AB	22	20	90.9%	2	9.1%
POT	5	0	0.0%	5	100.0%

one score, the average effect size across classrooms was calculated for these studies (studies 6, 14) first and then the descriptor was assigned according to the appropriate guidelines.

The graphic representation of the comparison between the SVA protocol scores and the reported effect sizes is presented in Figures 2 and 3, and data tables can be found in the

Appendix H. For DB, the SVA scores generally aligned well with the effect sizes reported in the reviewed studies, except for studies 13 and 14. Harry et al. (2023) (study 13) implemented MDB with four baselines. Simple visual analysis suggested the presence of the effect between baseline and intervention within each tier. However, the overlap between introduction of intervention and a baseline in adjacent tiers is only one session, which does not meet the staggered introduction of intervention criterium by Wolfe et al. (2019). Therefore, the suboptimal research design in this case prevented any conclusions regarding the positive effect of Tootling on DB. In study 14 (Morris, 2024) three data points presented in the withdrawal phase failed to demonstrate a trend, meaning the study did not meet the necessary criteria to conclude a functional relationship between Tootling and DB.

For AB, the SVA scores mostly aligned well with the effect sizes reported in the reviewed studies except for studies 6 and 13. Study 6 (Derieux, 2019) showed a low level of contrast between baseline and treatment phases, minimal immediacy of change, and high overlap between the baseline and treatment phases. These issues prevented the study from achieving a score above 3, which is required to conclude a functional relationship even though a moderate effect was reported by the author. Study 13 (Harry et al., 2023) demonstrated the same issue with the design of their AB evaluation as it did with DB evaluation. Therefore, the suboptimal research design of this study prevented any conclusions regarding the positive effect of Tootling on AB.

Social validity of Tootling in Classroom Level Studies

Social validity was assessed for teachers in 29 (all studies except study 14) of the total 30 classrooms and for students in 19 classrooms (studies 2, 4, 8 - 13 15, 16). The tools used for evaluation of social validity among teachers were Intervention Rating Profile (IRP-15) by Martens et al. (1985), Behaviour Intervention Rating Scale (BIRS) by Von Brock and Elliott (1987), User Rating Profile-Intervention (URP-I) and User Rating Profile- Intervention Revised

Figure 2.

Visual analysis score and effect sizes for DB across classrooms.

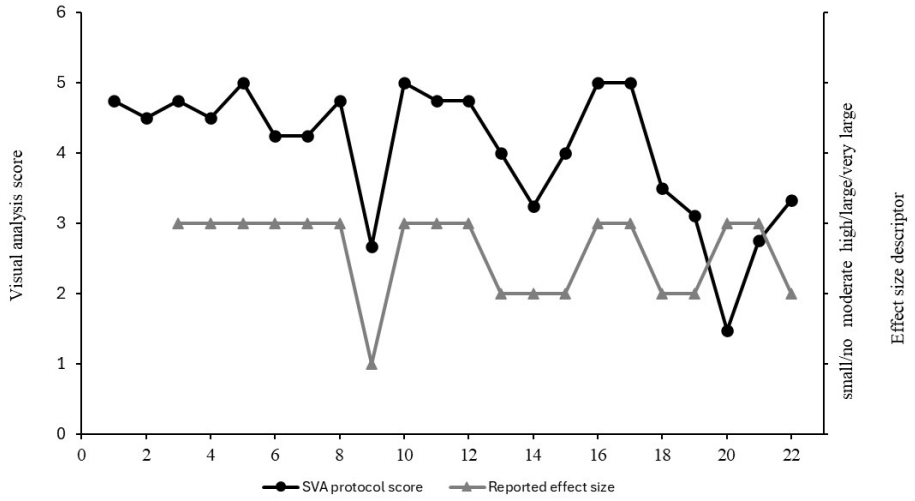
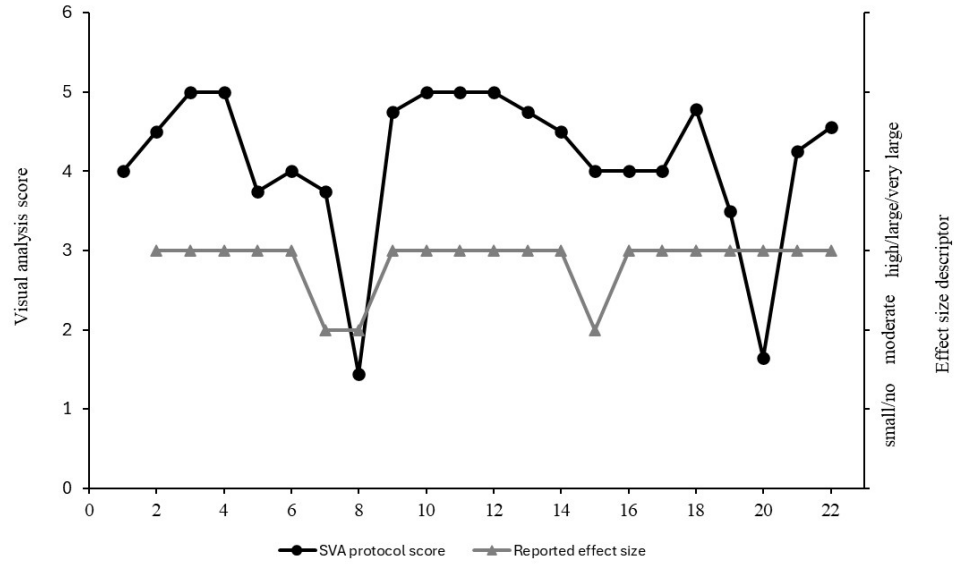


Figure 3.

Visual analysis scores and effect sizes for AB across classrooms.



(URP-IR) by Briesch et al. (2013). The BIRS was used in 16 classrooms (studies 5, 7, 8, 9, 11, 12, 13), the IRP-15 in 5 (studies 1, 2, 3, 4), and the URP-IR in 8 (studies 6, 10, 15, 16). Measures of social validity among students included Children's Intervention Rating Profile (CIRP) by Witt and Elliott (1985); Witt and Martens (1983), and Children's Usage Rating Profile (CURP) by Briesch and Chafouleas (2009). The CIRP was used in 14 classrooms (studies 2, 4, 8, 9, 11, 12, 13) and the CURP in 5 (studies 10 and 15). Table 9 summarises all the social validity measures used in classroom-level studies. All classrooms where social validity was evaluated reported that the intervention was socially valid and acceptable as judged by both teachers and students. Derieux (2019) (study 6) attempted to measure student social validity but was unable to collect enough responses from students in their 3 classrooms, so the student social validity data for those classrooms was not reported and they were not included in the analysis. Chaffee et al. (2020) and Tran (2024) (studies 10 and 15), both using URP-IR did not report a single score but presented average scores across all questions. Therefore, the overall mean was calculated according to the instructions by Chafouleas (2011).

Table 9.

Social validity assessment in classroom level studies.

Group	Total Classrooms Evaluated	Instrument Used	Classrooms Using Instrument	
			Classrooms	%
Teachers	29	BIRS	16	55.2%
		IRP-15	5	17.2%
		URP-IR	5	17.2%
		URP-I	3	10.3%
Students	19	CIRP	14	73.7%
		CURP	5	26.3%

The average levels of social validity were calculated across classrooms that used the same rating scales. For the Children's Intervention Rating Profile (CIRP), some studies reported mean scores, while others provided total scores. Wright (2019), Harry et al. (2023), and Enright (2023) (studies 9, 12, 13) reported both values. To calculate the average CIRP value, mean scores were multiplied by the number of questions in each study. Lum et al. (2019) (study 8) used a traditional questionnaire with seven questions, while McHugh et al. (2016) (study 4) implemented a modified CIRP with eight questions. The results presented in Table 10 demonstrate that all scores were well above 50% of their maximum, consistently reflecting acceptable levels of social validity and utility of Tootling.

Table 10.

Mean social validity scores in classroom level studies.

Target participants	Rating scale	Mean score	Max score	% of max score
Teachers	IRP-15	80.6	90	89.6%
	BIRS	4.7	6	78.1%
	UPR-IR	4.3	6	72.4%
Students	CIRP	31.9	42	75.9%
	CURP	3.4	4	84.0%

Studies with Student Level Data

Demographics of Student Level Studies

Among 18 selected studies, 3 studies presented data for individual students (studies 12, 17, 18). There were 12 participants in these studies aged from 7 to 17 years old attending grades 3 to 10. Six students attended general classrooms (studies 12, 18) and the other 6 students attended special classrooms (study 17). Among 12 students 11 were male (studies 17, 18) and 1 was female (study 12). There were 6 students with an ED (study 17), 1 student with ASD (study 12), 2 students with ADHD (study 18), and 2 students with a SLD (study

18). Two students did not have documented disabilities but were receiving Tier 2 support (study 18). One classroom with six students was also part of the school which implemented PBIS frameworks at the time of the study (study 17). Ten students were located in the US (studies 17, 18) and one student was in New Zealand (study 12). The demographic characteristics of participants in student-level studies are summarised in Table 11.

Research Characteristics of Student Level Studies

Among three studies presenting student-level data one was a peer-reviewed article (study 18) and two were non-peer-reviewed works (studies 12, 17). Studies were conducted by three independent research teams from three different institutions: the University of Texas (study

Table 11.

Demographic characteristics of students from student level studies.

Characteristic	Category	Studies, % of total		Students, % of total	
Total Studies	3 studies				
Total Participants	12 participants				
Grade Levels	Grades 3 to 10				
	Elementary/primary	3	100.0%	8	66.7%
	Middle	1	33.3%	2	16.7%
	High	1	33.3%	2	16.7%
Age Range	7 to 17 years old				
Gender	Male	2	66.7%	11	91.7%
	Female	1	33.3%	1	8.3%
Special Needs/Supports*	Emotional Disability (ED)	1	33.3%	6	50.0%
	Autism Spectrum Disorder (ASD)	1	33.3%	1	8.3%
	ADHD	1	33.3%	2	16.7%
	Specific Learning Disability (SLD)	1	33.3%	2	16.7%
	No documented disabilities, Tier 2 support	1	33.3%	2	16.7%
Using SWPBIS Framework	Yes	2	66.7%	6	50.0%
	No	1	33.3%	6	50.0%
Locations of Studies	US	2	66.7%	11	91.7%
	New Zealand	1	33.3%	1	8.3%

Note: *One student had two diagnoses

17), the University of Waikato (study 12), and the University of Georgia (study 18). Two of these institutions are located in the United States with 11 participating students (studies 17, 18), while one study with one participating student is in New Zealand (study 12). Two studies implemented withdrawal designs (studies 12, 17) and one study implemented an alternating treatments design (study 18). Two studies (studies 12 and 18) implemented only one Tootling variant while two variants, teacher-led and student-led, were implemented in Study 17. AB was measured in all three studies covering 12 students; DB were measured in two studies for 11 students (studies 17, 18); and POT was measure in one study for 5 students (study 18). Two studies (12 and 18) had a fading phase, and authors of one study with one student performed a follow-up (study 13). Initial screenings were conducted by the researchers of two studies (studies 17, 18). The length of observation sessions ranged from 20 minutes in 2 studies with 11 students (studies 17, 18) to 35 minutes in one study with one student (study 12). Two studies with 6 participants implemented momentary time sampling (studies 12, 18) and one study with another 6 participants implemented whole and partial interval recording (study 17). Social validity was measured in all three studies for their seven participating teachers. All three studies measured social validity for students, covering 11 students. And effect sizes were calculated in two studies with 7 students (studies 12, 17). Table 12 summarises research characteristics of student-level studies.

Tootling Procedure across Student Level Studies

Different studies implemented Tootling for different lengths of time during the day. While one study (study 17) implemented Tootling for the whole day with reporting permitted at any time throughout the day, another study (study 12) opted for half-a-day implementation with reports posted during the lunchbreak. A third study (study 18) implemented Tootling for the duration of one class with the posting of the reports at the end of that class. None of the research teams conducted a vote for the intervention name but all of them sought student

Table 12.

Research characteristics of student level studies.

Characteristic	Category	Studies (% of total)		Students (% of total)	
Publication Type	Peer-reviewed article	1	33.3%	5	41.7%
	Non-peer-reviewed work	2	66.7%	7	58.3%
Research Institutions	University of Texas	1	33.3%	6	50.0%
	University of Waikato	1	33.3%	1	8.3%
	University of Georgia	1	33.3%	5	41.7%
Research locations	US	2	66.7%	11	91.7%
	New Zealand	1	33.3%	1	8.3%
Research Teams	3 independent research teams				
Research Design	Withdrawal Design	2	66.7%	7	58.3%
	ATD	1	33.3%	5	41.7%
Tootling Variants	Single variant	2	66.7%	7	58.3%
	Two variants	1	33.3%	5	41.7%
Dependent Variables	AB	3	100.0%	12	100.0%
	DB	2	66.7%	11	91.7%
	POT	1	33.3%	5	41.7%
Fading	Yes	1	33.3%	5	41.7%
	No	2	66.7%	7	58.3%
Follow-Up	Yes	1	33.3%	1	8.3%
	No	2	66.7%	11	91.7%
Participant Screening	Yes	2	66.7%	11	91.7%
	No	1	33.3%	1	8.3%
Recording Procedures	Momentary Time Sampling	2	66.7%	6	50.0%
	Interval Recording	1	33.3%	6	50.0%
Observation Session Duration	20 minutes	2	66.7%	11	91.7%
	35 minutes	1	33.3%	1	8.3%
Social Validity	Teachers	3	100.0%	7	58.3%
	Students	3	100.0%	12	100.0%
Effect Size Calculated	Yes	2	66.7%	7	58.3%
	No	1	33.3%	5	41.7%
Treatment Integrity Measured	Yes	3	100.0%	12	100.0%
	No	0	0.0%	0	0.0%

input in choosing the rewards. One classroom (study 17) used technology (Padlet, Google sheet) to aid their Tootling procedure. Two studies involving 11 students did not have allocated time or limits for posting tootles (studies 17, 18). However, in one of those classrooms each student was assigned a “shoutout partner” and their Tootling was limited to that partner-student. One study with one student had a limit of three reports per day (study 12). The format of reports varied across studies and students. One study used a traditional

free-text format for all their 6 students (study 17), one study used a standardised format for their one student (study 12), and one study used a standardised format for two of their students and a free-text format for another two (study 18). One study (study 17) did not report intended time for reaching the goal by the students. In another study with one participant (study 12) a goal could be reached within one day but needed to be achieved for three days to receive a reward. In a third study with 5 students (study 18), goals were calculated to be reached approximately twice a week. The number of tootles in the goal was also different across the studies. In one study (study 17) the initial goal was set as the number of students in the classroom multiplied by three and then increased. In another study (study 12) a goal was chosen randomly every day and was ranging from 1 to the number of students present that day. In a third study, goals were kept consistent at 40 tootles for two students with standardised format of Tootling cards and 42 for another three (study 18). In two studies covering 11 students where goals were set, those goals were known to students (studies 17, 18). In a study with randomised goals (study 12), the goal was kept secret till the end of the daily session. Two studies used “Total tootles” goal criteria (studies 17, 18) and in the third study (study 12) criteria were tied to the number of students who received tootles during the session. All three studies provided feedback to their students at the end of the session with two studies traditionally reading aloud five tootles (studies 12, 18) and in one study all daily tootles were displayed on the screen at the end of the day (study 17). All three studies utilised interdependent group contingencies. In two studies with 7 students Tootling was implemented by the teachers (studies 12, 17) and in one study with 5 students, teacher- and student-led conditions were utilised (study 18). The characteristics of Tootling procedure in student-level studies can be found in Table 13.

Table 13.

Tootling procedure in student level studies.

Characteristic	Category	Number of Studies, % of total		Number of students, % of total	
Implementation Duration	Whole day	1	33.3%	6	50.0%
	Half-day	1	33.3%	1	8.3%
	One class duration	1	33.3%	5	41.7%
Timing of Tootling Submissions	Any time during school day	1	33.3%	6	50.0%
	50-60 minutes during lunchtime	1	33.3%	1	8.3%
	At the end of the class	1	33.3%	5	41.7%
Submission Limits	No limits	2	66.7%	11	91.7%
	Limit of three reports per day	1	33.3%	1	8.3%
Use of Technology	Yes	1	33.3%	6	50.0%
	No	2	66.7%	6	50.0%
Tootling Card Format	Free-text format	2	66.7%	8	66.7%
	Standardised format	2	66.7%	3	25.0%
Group Contingency Type	Interdependent	3	100.0%	12	100.0%
Voting for Intervention Name	No	3	100.0%	12	100.0%
Student Input in Choosing Rewards	Yes	3	100.0%	12	100.0%
Goal Setting Strategy	Not reported	1	33.3%	6	50.0%
	Goals achievable within one day (three-day reward criteria)	1	33.3%	1	8.3%
	Goals achievable twice per week	1	33.3%	5	41.7%
Goal Criteria	Total tootles	2	66.7%	11	91.7%
	Tootles received per student	1	33.3%	1	8.3%
Goal Transparency	Known to students	2	66.7%	11	91.7%
	Kept secret	1	33.3%	1	8.3%
Public Acknowledgment of Reports	Read aloud five tootles	2	66.7%	6	50.0%
	Displayed all tootles on screen	1	33.3%	6	50.0%
Public Posting of Progress	Yes	3	100.0%	12	100.0%
Tootling Implementers	Teacher	3	100.0%	12	100.0%
	Teacher and students	1	33.3%	5	41.7%

Findings Across Student Level Studies

DB was measured in two studies across 11 students (studies 17, 18) and demonstrated a positive effect in all of them. One study involving 6 students (study 17) reported Tau-U effect sizes for DB. All the effects sizes in this study were in a "very large" range. AB was measured in three studies covering all twelve students and was also found effective for all of them. Two studies involving 7 students (study 12, 17) reported Tau-U effect sizes, all of which were also in the "very large" range. One study measured POT across their 5 students (study 18) and a functional relationship between Tootling and POT was not established for any of those students. Effectiveness of Tootling according to reported findings in student-level studies is summarised in Table 14.

Systematic Visual Analysis of Student Level Studies

Consistent with the studies using classroom-level analysis, systematic visual analysis was conducted for those employing student-level analysis. Of the three studies implementing student-level analysis, only two, involving seven students and both using withdrawal designs,

Table 14.

Effectiveness of tootling across target behaviours in student level studies.

Target Behaviour	Students Evaluated	Effective in Students	Effect Size reported	High/Large/Very Large Effect	Moderate Effect	Small/No Effect
DB	11	11 (100.0%)	6	6 (100.0%)	0 (0.0%)	0 (0.0%)
AB	12	12 (100.0%)	7	7 (100.0%)	0 (0.0%)	0 (0.0%)
POT	5	0 (0.0%)	0	0 (0.0%)	0 (0.0%)	5 (100.0%)

were suitable for the SVA protocol (studies 12, 17). DB was measured in one study and achieved the maximum score of 5 for all participants (study 17). AB was measured in both studies also receiving scores of 5 for all participants. None of the two studies measured POT. The results of SVA across studies and students are presented in Table 15.

Table 15.

Effectiveness of Tootling based on systematic visual analysis score in student level data.

Target Behaviour	Total Students Evaluated	Effect Demonstrated		Effect Not Demonstrated	
		Students	%	Students	%
DB	6	6	100.0%	0	0.0%
AB	7	7	100.0%	0	0.0%

Figure 4 presents visual analysis scores and effect sizes for DB, while Figure 5 displays the same for AB. The SVA scores aligned well with the effect sizes reported in the reviewed studies for both DB and AB. None of the student-level studies evaluated with SVA were found implementing suboptimal research designs therefore all the studies are included in further summary. However, the quality of the evidence presented in study 18 could not be evaluated by the SVA because of the research design of this study.

Social Validity of Tootling in Student Level Studies

Social validity (SV) was assessed for both teachers and students across all three studies, involving 7 teachers and 12 students. For teachers, the BIRS was used in two studies with a total of four teachers (studies 12, 17); in a third study a custom measure was developed by its authors and administered it to their three teachers (study 18). The BIRS scores indicated moderate to high social validity. The custom measure reported high acceptability and social validity, though it did not provide specific scores.

For students, each study used a different tool to assess social validity. Two studies, involving 11 students (study 17, 18), used custom questionnaires, which did not provide specific scores but generally indicated a positive attitude toward Tootling. The third study, involving one student, used the CIRP (study 12). The CIRP score in this study averaged data from one student with special needs and other non-SN students who participated in the study but were not included in this review. While this approach limits the accuracy of capturing the target

Figure 4.

Visual analysis scores and effect sizes for DB across students.

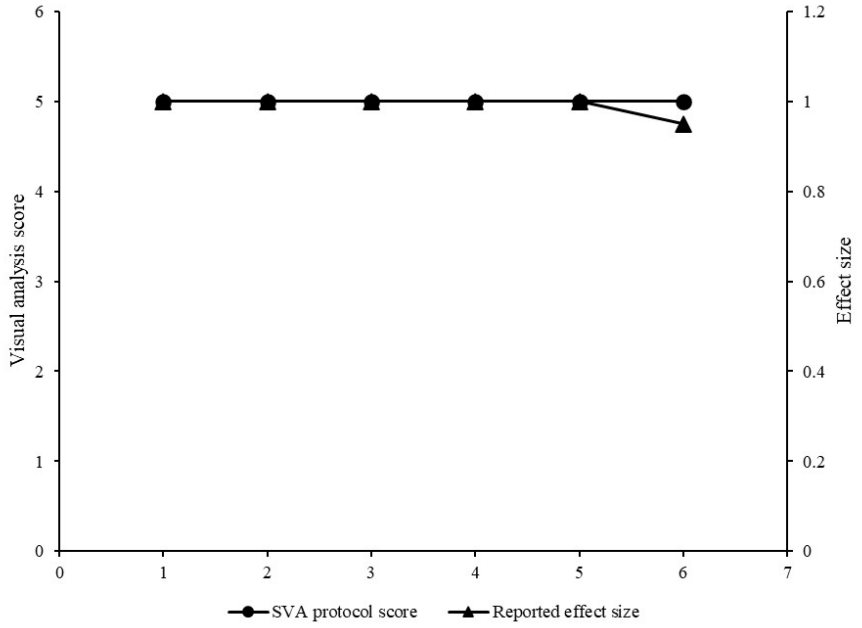
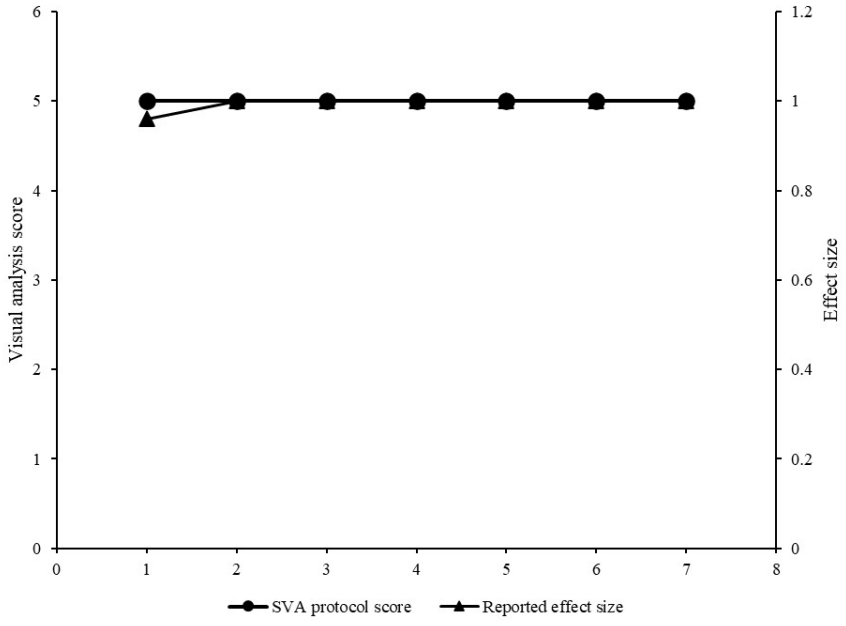


Figure 5.

Visual analysis scores and effect sizes for AB across students.



student's perspective, the average score suggests moderate social validity for Tootling and has been included in this review. Only one study followed up to determine whether the teacher continued using Tootling after the researchers had withdrawn (study 12), and the teacher did not continue, despite the reported moderate social validity of Tootling. Table 16 provides a summary of social validity evaluation in student-level studies.

Table 16.

Social validity assessment in student level studies.

Measure	Studies		Participants	
	Total	**SV reported (%)**	**Total**	**SV reported (%)**
Teacher social validity	3	3 (100.0%)	7	7 (100.0%)
Student social validity	3	3 (100.0%)	12	12 (100.0%)

Discussion

The aim of this study was to evaluate the effectiveness of Tootling for general education classrooms, that included children with special needs, and children with special needs specifically; and explore Tootling as a strategy supporting inclusion. Twenty studies were initially selected for this review. Following an evaluation against What Works Clearinghouse quality standards, one study was excluded completely and three studies partially. For studies that met the quality assessment criteria, an analysis was conducted to examine and compare various aspects, including demographics, methodologies, Tootling procedures, and research findings. Additionally, a systematic visual analysis (Wolfe et al., 2019) was performed for studies utilizing withdrawal and multiple baseline designs. Sixteen studies with 30 participating classrooms presented classroom-level data and 3 studies presented student-level data for a total of 12 students. All classrooms in this review included at least one student with special needs, and all students from the student-level studies had either a diagnosed disability or individual support assigned to them at the time of the study. The data generated from this analysis forms the basis for addressing this study's research questions.

Research Questions

1. **Is Tootling effective in improving classroom behaviours in general education classrooms that include students with special needs?**

The effectiveness of Tootling, according to the reported effects and effect sizes, varied across target behaviours in the reviewed classrooms. Tootling was found to be highly effective in addressing AB and DB. It was effective in reducing disruptive behaviour in 86.21% of classrooms, with 58.62% demonstrating large or very large effect sizes. Even greater effectiveness was observed in increasing appropriate behaviour, with Tootling being effective in 96.43% of classrooms and 78.57% showing large or very large effect sizes.

Although Tootling was reported to be effective in decreasing passive off-task behaviour in 75.00% of classrooms where it was measured, the effect was generally moderate. Given the small number of classrooms where POT was measured, it can be concluded that while Tootling may be effective for POT, it is less effective compared to its impact on DB and AB, and further research is needed. In summary, the Tootling procedure was highly effective for decreasing disruptive behaviour and increasing appropriate behaviour in the reviewed classrooms according to the data reported in the reviewed studies.

Systematic visual analysis demonstrated similar results with Tootling being effective in 86.36% of classrooms for DB and 90.91% of the classrooms for DB while being ineffective for POT. The scores from systematic visual analysis aligned well with the effect sizes reported in the reviewed studies (Figure 2) except for two studies. Morris (2024) reported large effect sizes; however, three data points in the withdrawal phase did not demonstrate a trend. This did not allow this study to receive a sufficient systematic visual analysis score. Similarly, although large effect sizes were reported by Harry et al. (2023), the flaws in research design prevented it from receiving a high SVA score. Overall, the agreement between the reported and independently acquired data was demonstrated.

Three studies did not demonstrate Tootling effect for DB: Derieux (2019), Morris (2024), and Harry et al. (2023); only two studies did not demonstrate an effect for AB: Derieux (2019) and Harry et al. (2023). Derieux (2019) employed an alternating treatments design, switching treatments every session. Three conditions were alternated in random order: a traditional Tootling treatment, a comparison writing treatment, and a no-treatment control. Three classrooms were combined under a multiple baseline design in this study. According to classroom effect sizes, Tootling was effective in Classroom A, produced mixed effects in Classroom B, and was not effective in Classroom C. Derieux (2019) identified several limitations, including confusion among teachers and students regarding the intervention being

implemented on each day, which likely blurred the distinctions between conditions and affected the results. Additionally, frustration caused by this confusion may have hindered improvements in student behaviour (Derieux, 2019). Another study using an alternating treatments design, Tran (2023), presented Tootling in all conditions but varied the reward criteria: based on the number of students writing tootles or the number of students receiving tootles. Therefore, the alternating procedures were both Tootling procedures with minimal difference. Tran (2023) reported large effect sizes for both DB and AB. A comparison of these two studies suggests that to achieve significant effects, Tootling should be implemented consistently over multiple sessions.

In Morris (2024) visual analysis shows clear effect during the first phase change but the withdrawal phase was not stable enough to allow for the second phase change effect to be seen. Therefore, Tootling was likely effective, but the design of the study does not allow for robust replication of the effect.

Interestingly, there was considerable variation in Tootling procedures across studies. For instance, some studies implemented Tootling throughout the entire day with an unlimited number of reports (e.g., Cihak, 2009), whereas others limited the intervention to a single teaching block or a lesson as short as 30 minutes. Some studies restricted the number of reports, while others limited the time allocated for reporting. Although, most studies used traditional reporting with physical cards, some studies used technology, such as Class Dojo, to assist the Tootling process. Some variations lacked important elements of traditional Tootling. For example, Harry et al. (2023) implemented two variations of Tootling. The first variation did not include feedback in the form of reading five tootles aloud at the end of the day, while the second one included posting of all students' reports for the day on a special board. Even though an important component of traditional Tootling was removed in the first variation, it still was highly effective both according to the authors and systematic visual

analysis conducted by this author. The second variation showed further improvement although, according to the authors, marginal.

Overall, this review found Tootling effective in decreasing DB and POT and increasing AB in the classrooms with students with special needs which is consistent with the research in all general classrooms.

2. **Is Tootling an evidence-based strategy for enhancing classroom behaviours in general education settings that include students with special needs?**

According to WWC (Institute of Education Studies, 2017) recommendations for combining studies, a minimum of five studies performed by three different research teams in different locations with at least 20 single-case examples must be present for the intervention to be considered evidence-based. For the current review, the results of 16 studies across 30 target classrooms were evaluated. Those studies were performed by 6 independent teams across different locations. Therefore, the "3-5-20" threshold for the evidence-based practice is fulfilled and Tootling can be considered an evidence-based strategy for enhancing classroom behaviours in general education settings that include students with special needs.

Additionally, classrooms from the reviewed studies included on average four students with special needs and covered all grade levels, including both PBIS and non-PBIS schools. A wide range of special needs was covered in the studies, ranging from physical disabilities to ASD and ADHD diagnoses to emotional disturbances. Although most studies were conducted in various locations in the US, studies in New Zealand also showed effectiveness; therefore, increasing the generalisability of the findings across cultural contexts.

3. **Is Tootling effective in improving classroom behaviours specifically for students with special needs?**

Only three studies included in this analysis evaluated the effect of Tootling on students' behaviour at the level of the individual student. Two of them measured DB across 11 students and a functional relationship was established for all of them. Only one study with 6 students reported effect sizes and all effect sizes were very large. All 3 studies evaluated the effect of Tootling on AB across 12 students and, again, found positive effects for all of them. Effect sizes were reported for 8 students and all of them were very large. This demonstrates high effectiveness of Tootling in improving disruptive and appropriate/academic behaviours. One study evaluated the effect of Tootling on POT across 5 students and found no evidence of a functional relationship. Systematic visual analysis was conducted for two studies and supported the conclusions reported in those studies, demonstrating high scores for DB and AB for Enright (2023) and Salinas (2021).

Despite the small number of studies and participants, there were variations in Tootling procedures across the studies. Thus, the duration of Tootling session varied from a 30-minute session to a whole day session. There were studies with and without the use of technology, reporting limits, set and randomised tooling goals, different goal criteria and goal transparency. Additionally, all grade levels were presented in the study and a variety of special needs. The intervention was perceived socially valid and acceptable by both teachers and students.

Overall, these studies demonstrated high effectiveness of Tootling in improving classroom behaviours specifically for students with special needs across the variety of Tootling variations, student needs and ages.

4. **Is Tootling an evidence-based strategy for improving classroom behaviours among students with special needs?**

Only a small number of studies reporting student-level data met the criteria for this review. Overall, there were three studies conducted by three independent research teams

across 12 participants. Therefore, the “3-5-20” threshold for the evidence-based practice is not fulfilled. Although these studies showed very promising results, more research is needed where the behaviour of students with special needs is measured at the level of the individual student.

Overall Conclusions

This review concludes that Tootling is an effective intervention for improving class-wide student behaviour, which is consistent with the findings of Collins et al. (2020) in their meta-analysis of peer-reporting interventions. Moreover, Tootling showed promising results in enhancing the behaviour of students with special needs when measured individually, suggesting a potentially greater than average impact on these students. Reducing disruptive and off-task behaviours while increasing appropriate and academic behaviours creates more time for instruction, potentially leading to improved academic achievement. This enhanced instructional time can enable teachers to better support individual students, providing students with special needs increased opportunities for assistance. Thus, Tootling appears to support academic outcomes for all students, with particular benefits for those with special needs. However, this conclusion should be taken with caution due to a limited number of studies which presented student-level data for students with special needs.

Tootling is a practical and economical intervention that does not require removing students from the classroom or providing individual support. In classrooms where the average number of students with special needs exceeds four, relying on individual behaviour programmes would demand significant resources and specialist involvement. By contrast, Tootling is a low-effort, teacher-led strategy that can be partially student-driven, as demonstrated by Wright (2019) and Thoele (2024). Thoele's study highlighted that their students overwhelmingly preferred a student-led approach, which further reduced teacher

workload. This classwide intervention mitigates the need for more resource-intensive, individualised strategies.

Additionally, Tootling promotes social inclusion by avoiding the stigmatisation of students with behavioural challenges. Since it is implemented across the entire class, individual students are not singled out, helping to foster a sense of belonging and inclusion while reducing the risk of social isolation. Tootling can also be adapted to further support inclusion. For example, goal criteria can focus on the number of students receiving recognition rather than the total number of tootles, as seen in Lum et al. (2019), Enright (2023), and Thoele (2024). This adjustment ensures that every student's positive behaviour is acknowledged, reinforcing social cohesion and maintaining a focus on the positive.

The classwide implementation of Tootling aligns well with the Response to Intervention (RTI) and Positive Behavioural Interventions and Supports (PBIS) frameworks, particularly as a Tier 1 support strategy in classrooms with students with special needs. By addressing behaviour at the whole-class level, Tootling avoids unnecessary labelling and provides a proactive approach to managing behaviour. This review demonstrated the successful application of Tootling across grade levels and various special needs. Its flexibility allows for adjustments to fit the unique needs of each classroom or school, making it suitable for integration into a school-wide PBIS framework.

Tootling also provides opportunities for developing social skills. While research on its effects on pro-social behaviour is limited and inconsistent, studies by Crewdson et al. (2022) and Wright et al. (2022) indicated that Tootling can enhance specific pro-social behaviours in at-risk children following social skills training (SST). By offering students opportunities to practice these newly acquired skills and receive peer reinforcement, Tootling may be particularly beneficial for students with special needs when combined with targeted social

skills training. Additionally, social skills that are targeted in SST plus Tootling could include those that encourage typical children to socially include their SN peers.

In conclusion, Tootling holds significant potential as a powerful inclusion strategy in general education classrooms. It not only addresses behavioural challenges but also promotes social inclusion and skill development, supporting the diverse needs of students. However, further research is needed to explore its long-term effects and optimise its implementation for students with special needs.

Implications for Practice

Studies in this review demonstrated a wide variety of ways that Toolting can be flexible to fit in different educational environment yet remain effective. Thus, full-day sessions can be implemented in elementary classrooms, while for higher grade levels, Tootling can be implemented during one block or class. Some variations are particularly useful for inclusive classrooms. For example, a standardised format of Tootling cards was implemented in several studies. Such a format may be useful not only for students whose writing skills are still developing but also for students with physical difficulties.

Thoele (2024) introduced using a “shout out partner” for their students in a special classroom. In their approach, each student was assigned a partner, and the students could only tootle on their assigned partner. This arrangement encouraged students to observe more closely the behaviour of a particular student and to ensure that every student gets their good behaviour noticed and reported. It was reported, however, that students expressed such arrangement to being boring and would prefer to be able to tootle any students from the class. Therefore, a better strategy for involving all children could be to set goal criteria based on the number of children receiving tootles. Another observation from this study was that Tootling is less interesting in a small special classroom, and children would prefer to play this game in

big classrooms. This emphasises the importance of including SN students in general classrooms to enrich their social lives.

Limitations and Future Research

This review aimed to evaluate the feasibility and effectiveness of Tootling as an intervention supporting inclusion in general education classrooms. It focused on studies where Tootling was the sole intervention implemented in classrooms with at least one student with special needs. The findings, reported by the original authors and validated through independent systematic visual analysis, demonstrates strong evidence for the effectiveness of Tootling in increasing appropriate behaviour (AB) and reducing disruptive behaviour (DB) in general education classrooms with special needs students. However, several limitations must be acknowledged.

The first limitation is that a systematic visual analysis could not be conducted for studies using an alternating treatments design. As the protocol implemented in this study was developed only for withdrawal and MB design studies, an independent assessment for ATD studies was not possible. Therefore, future studies could use other methods of independent assessment of the reviewed studies' results, such as conducting a meta-analysis. Nevertheless, the small number of ATD studies included in this review is unlikely to significantly affect the overall conclusions. Additionally, the findings of all but two studies were consistent with the systematic visual analysis, enhancing confidence in the results reported by the authors.

Another limitation is the small number of studies with student-level data. While these studies demonstrated promising results, they do not offer robust evidence for Tootling's effectiveness specifically for students with special needs. Moreover, the demographic characteristics of the participants in these studies lack diversity. For instance, 10 of the 11 individual participants were male, and most were from upper elementary classrooms. Younger students were not represented, and other grade levels were under-represented. Future

research should address these gaps by focusing on individual students across a broader range of demographic factors, including age, gender, and cultural backgrounds.

A third limitation is that classrooms at the middle school level are underrepresented in the reviewed studies. Of the 30 classrooms examined, only one was from a middle school, while elementary and high school classrooms were represented more evenly. Middle school can be a particularly challenging period due to the increasing complexity of social interactions and the heightened importance of academic achievement. Research confirms that behaviour challenges are more common in middle than elementary schools (Kaufman et al., 2010). Effective behaviour management during this stage could play a crucial role in setting the foundation for success in high school. Therefore, more research in middle school classrooms is needed.

The overall small number of studies on Tootling which include students with special needs required including non-peer reviewed works in the review. This could affect the overall quality of the research literature for this review. Therefore, as the research base on Tootling grows, future researchers may be able to synthesise exclusively peer-reviewed studies.

A limitation may be present in the fact that, because of the differences in effect size reporting, it was impossible to extract and compare effect sizes from the first phase change across all classrooms. Therefore, although mostly clear first phase change effect sizes were taken for classrooms in MDB, the effect sizes from withdrawal studies were average across all phases which compromised consistency. There is a possibility that the effect from a second phase change may either decrease or increase the average value. For example, if the intervention is so effective that it produced a carry-over effect from the first intervention to the withdrawal phase, the level of target behaviour may not return to the initial baseline levels reducing the contrast between withdrawal and intervention phases. Therefore, even though such effects increase researchers' confidence in the effectiveness of intervention, the opposite

is reflected in the average effect size. Therefore, future researchers may explore the possibilities of more consistent comparison between participants (classrooms and students).

The lack of maintenance procedures, identified in the overview of Tootling literature, was also evident in the reviewed studies. Research indicates that Tootling, as an antecedent intervention, requires continuous and consistent implementation to sustain its effects. However, follow-up assessments, though infrequent, reveal that teachers rarely continue using Tootling once the research team withdraws. This happens despite positive evaluations of the intervention through social validity measures. This situation highlights the need to design and implement structured maintenance procedures with long-term follow-ups and to develop strategies that incentivise teachers to sustain the implementation of Tootling in their classrooms. If Tootling is established as a recommended Tier 1 strategy and promoted at the school or even systems level, its continued and consistent implementation becomes more likely. Such widespread adoption aligns with the concept of Tootling as an inclusive strategy, fostering long-term behavioural benefits for both typically developing students and those with special needs.

In conclusion, to establish Tootling as an evidence-based strategy for promoting inclusion, future research should focus on collecting more student-level data for students with special needs while ensuring greater demographic diversity across grade levels, genders, and cultural contexts for both classroom and student-level data. Additionally, there is a need for the design and implementation of maintenance procedures and long-term follow-ups for both classroom and student-level data collection.

References

Alberto, P., & Troutman, A. C. (2013). Applied behavior analysis for teachers.

Artman-Meeker, K. M., & Hemmeter, M. L. (2013). Effects of Training and Feedback on Teachers' Use of Classroom Preventive Practices. *Topics in Early Childhood Special Education*, *33*(2), 112-123. https://doi.org/10.1177/0271121412447115

Babyak, A. E., Luze, G. J., & Kamps, D. M. (2000). The good student game: Behavior management for diverse classrooms. *Intervention in school and clinic*, *35*(4), 216-223.

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of applied behavior analysis*, *2*(2), 119-124.

Barth, J. M., Dunlap, S. T., Dane, H., Lochman, J. E., & Wells, K. C. (2004). Classroom environment influences on aggression, peer relations, and academic focus. *Journal of school psychology*, *42*(2), 115-133.

Barton, E. E., Lloyd, B. P., Spriggs, A. D., & Gast, D. L. (2018). Visual analysis of graphic data. In *Single case research methodology* (pp. 179-214). Routledge.

Bear, G., & Manning, M. (2008). Best practices in classroom discipline. *Best practices in school psychology V*, 1403-1420.

Bear, G. G., Slaughter, J. C., Mantz, L. S., & Farley-Ripple, E. (2017). Rewards, praise, and punitive consequences: Relations with intrinsic and extrinsic motivation. *Teaching and Teacher Education*, *65*, 10-20. https://doi.org/https://doi.org/10.1016/j.tate.2017.03.001

Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the Good Behavior Game: A meta-analysis of single-case research. *Journal of positive behavior interventions*, *18*(3), 180-190.

Boyd, S., & Felgate, R. (2015). *" A Positive Culture of Support": Final Report from the Evaluation of PB4L School-Wide*. Ministry of Education Wellington, New Zealand.

Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal of behavioral education*, *17*, 4-23.

Briesch, A., & Chafouleas, S. (2009). Children's usage rating profile (actual). *Storrs, CT: University of Connecticut*.

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of school psychology*, *51*(1), 81-96.

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of educational research*, *51*(1), 5-32.

Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior modification*, *30*(5), 531-563.

Browne, K. (2013). Challenging behaviour in secondary school students : classroom strategies for increasing positive behaviour. *New Zealand journal of teachers' work*, *10*(1), 125-147.

Cashwell, T. H., Skinner, C. H., & Smith, E. S. (2001). Increasing Second-Grade Students' Reports of Peers' Prosocial Behaviors Via Direct Instruction, Group Reinforcement, and Progress Feedback: A Replication and Extension. *Education & treatment of children*, *24*(2), 161-175.

Cengher, M., Budd, A., Farrell, N., & Fienup, D. M. (2018). A Review of Prompt-Fading Procedures: Implications for Effective and Efficient Skill Acquisition. *Journal of Developmental and Physical Disabilities*, *30*(2), 155-173. https://doi.org/10.1007/s10882-017-9575-8

Center on PBIS. (2025). *What is PBIS?* https://www.pbis.org/pbis/what-is-pbis

Chaffee, R. K., Briesch, A. M., Volpe, R. J., Johnson, A. H., & Dudley, L. (2020). Effects of a Class-Wide Positive Peer Reporting Intervention on Middle School Student Behavior. *Behavioral disorders*, *45*(4), 224-237. https://doi.org/10.1177/0198742919881112

Chafouleas, S. M., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. . (2011). *Usage Rating Profile – Intervention (Revised)*. Storrs, CT: University of Connecticut. https://urp.uconn.edu/forms/

Choundira, R. (2022). *Tootling through a cultural lens: Effects of tootling on student and teacher behaviours in an inclusive school in Aotearoa* The University of Waikato].

Church, J. (2003). *The Definition, Diagnosis and Treatment of Children and Youth with Severe Behaviour Difficulties: A Review of the Research: Report Prepared for the Ministry of Education*. Ministry of Education.

Cihak, D. F., Kirk, E. R., & Boon, R. T. (2009). Effects of Classwide Positive Peer "Tootling" to Reduce the Disruptive Classroom Behaviors of Elementary Students with and

without Disabilities. *Journal of behavioral education*, *18*(4), 267-278. https://doi.org/10.1007/s10864-009-9091-8

Coie, J. D., & Cillessen, A. H. (1993). Peer rejection: Origins and effects on children's development. *Current directions in psychological science*, *2*(3), 89-93.

Collins, T. A., Drevon, D. D., Brown, A. M., Villarreal, J. N., Newman, C. L., & Endres, B. (2020). Say something nice: A meta-analytic review of peer reporting interventions. *Journal of school psychology*, *83*, 89-103. https://doi.org/10.1016/j.jsp.2020.10.002

Collins, T. A., Meredith Murphy, J., & Heidelburg, K. (2018). Promoting Supportive Peer Relationships Using Peer Reporting Interventions. In R. H. L. Nabors (Ed.), *Promoting prosocial behaviors in children through games and play: Making social emotional learning fun* (pp. 63-83). Nova Science Publishers Incorporated.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (Third, global edition. ed.). Pearson.

Crewdson, M., Skinner, C. H., Wilson, C., Fowler, K., Richardson, R., & Wilhoit, B. (2023). Enhancing first-grade students' social skill performance: The interaction of tootling and activities on social skill engagement. *Psychology in the schools*, *60*(10), 4061-4078. https://doi.org/10.1002/pits.22984

Crewdson, M. A., Richardson, R. D., Fowler, K., Skinner, C. H., Wright, S., & Cihak, D. (2024). Supplementing Social Skills Training With Tootling to Simultaneously Enhance First-Grade Students' Performance of Two Social Skills. *School psychology review*, *53*(3), 236-250. https://doi.org/10.1080/2372966X.2022.2043127

Crouch, P. L., Gresham, F. M., & Wright, W. R. (1985). Interdependent and independent group contingencies with immediate and delayed reinforcement for controlling classroom behavior. *Journal of school psychology*, *23*(2), 177-187.

Davidson, N. (1990). Small-group cooperative learning in mathematics. *Teaching and learning mathematics in the*, 52-61.

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of applied behavior analysis*, *12*(4), 573-579.

Derieux, J. (2019). *Is the Written Component of Tootling Effective? A Comparison of the Group Contingency to a Comparison Writing Procedure* [Master's Thesis, The University of Southern Mississippi].

Dillon, M. B. M., Radley, K. C., Tingstrom, D. H., Dart, E. H., & Barry, C. T. (2019). The Effects of Tootling via ClassDojo on Student Behavior in Elementary Classrooms.

School psychology review, *48*(1), 18-30. https://doi.org/10.17105/SPR-2017-0090.V48-1

Drabman, R. S., & Lahey, B. B. (1974). Feedback in Classroom Behavior Modification: Effects on the Target and Her Classmates. *Journal of applied behavior analysis*, *7*(4), 591-598. https://doi.org/10.1901/jaba.1974.7-591

Education Review Office. (2022). *Thriving at School? Education for Disabled Learners in Schools*.

Ennis, R. P., Royer, D. J., Lane, K. L., & Dunlap, K. D. (2020). Behavior-specific praise in pre-K–12 settings: Mapping the 50-year knowledge base. *Behavioral disorders*, *45*(3), 131-147.

Enright, C. J. (2023). *An investigation into the effects of tootling for instruction- following and on-task student behaviour in a New Zealand primary school* [Master's Thesis, The University of Waikato].

Ervin, R. A., Miller, P. M., & Friman, P. C. (1996). Feed the Hungry Bee: Using Positive Peer Reports to Improve the Social Interactions and Acceptance of a Socially Rejected Girl in Residential Care. *Journal of applied behavior analysis*, *29*(2), 251-253. https://doi.org/10.1901/jaba.1996.29-251

Fimian, M. (1988). *Teacher stress inventory*. Clinical psychology publishing.

Fisch, G. S. (1998). Visual inspection of data revisited: Do the eyes still have it? *The Behavior Analyst*, *21*, 111-123.

Gillies, R. M. (2016). Cooperative learning: Review of research and practice. *Australian Journal of Teacher Education (Online)*, *41*(3), 39-54.

Goss, A. L. (2019). *The Effects of Tootling Without External Rewards On Whole-Class Lessons in a Lower Elementary Classroom* [Master's Thesis, St. Catherine University].

Greer-Chase, M., Rhodes, W. A., & Kellam, S. G. (2002). Why the prevention of aggressive disruptive behaviors in middle school must begin in elementary school. *The clearing house*, *75*(5), 242-245.

Gresham, F. M. (2002). Teaching social skills to high-risk children and youth: Preventive and remedial strategies.

Gresham, F. M., & Gresham, G. N. (1982). Interdependent, dependent, and independent group contingencies for controlling disruptive behavior. *The Journal of Special Education*, *16*(1), 101-110.

Gresham, F. M., Sugai, G., & Horner, R. H. (2001). Interpreting outcomes of social skills training for students with high-incidence disabilities. *Exceptional children*, *67*(3), 331-344.

Grieger, T., Kauffman, J. M., & Grieger, R. M. (1976). Effects of peer reporting on cooperative play and aggression of kindergarten children. *Journal of school psychology*, *14*(4), 307-313. https://doi.org/10.1016/0022-4405(76)90027-3

Harris, V. W., & Sherman, J. A. (1973). Use and analysis of the "Good Behavior Game" to reduce disruptive classroom behavior 1. *Journal of applied behavior analysis*, *6*(3), 405-417.

Harry, S. W., Tingstrom, D. H., Dufrene, B. A., Dart, E. H., Radley, K. C., Lum, J. D. K., & Bates-Brantley, K. E. (2023). The Effects of Tootling Combined with Public Posting in High School Classrooms. *Journal of behavioral education*, *32*(3), 565-584. https://doi.org/10.1007/s10864-021-09462-1

Haydon, T., Kennedy, A., Murphy, M., & Boone, J. (2023). Positive peer reporting for middle school students with emotional and behavioral disorders. *Intervention in school and clinic*, *58*(4), 273-279.

Hayes, L. A. (1976). The use of group contingencies for behavioral control: a review. *Psychological Bulletin*, *83*(4), 628.

Henington, C., & Skinner, C. H. (1998). Peer monitoring. In *Peer-assisted learning* (pp. 237-254). Routledge.

Henricsson, L., & Rydell, A. M. (2006). Children with behaviour problems: The influence of social competence and social relations on problem stability, school achievement and peer acceptance across the first six years of school. *Infant and child development*, *15*(4), 347-366.

Horn, J. L. V. (2004). *Utility of Positive Peer Reporting to Improve Interactions Among Children in Foster Care*.[University of South Florida]. USF Tampa Graduate Theses and Dissertations.

Horner, R. H., & Sugai, G. (2015). School-wide PBIS: An example of applied behavior analysis implemented at a scale of social importance. *Behavior analysis in practice*, *8*, 80-85.

Institute of Education Science. (2025). *What Works ClearingHouse. About Us: What We Do*. https://ies.ed.gov/ncee/WWC/

Institute of Education Studies. (2017). *What Works Clearinghouse procedures and standards handbook version 4.0*. US Department of Education

Institute of Education Studies. (2020). *What Works Clearinghouse procedures and standards handbook, version 4.1*. US Department of Education

Institute of Education Studies. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. US Department of Education

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Applied Behavior Analysis.

Johansen, A., Little, S. G., & Akin-Little, A. (2011). An examination of New Zealand teachers' attributions and perceptions of behaviour, classroom management, and the level of formal teacher training received in behaviour management. *Kairaranga*, *12*(2), 3-12. https://doi.org/10.54322/kairaranga.v12i2.164

Johnson, D. W., & Johnson, R. T. (1992). Implementing cooperative learning. *Contemporary Education*, *63*(3), 173.

Johnston, J. M., Pennypacker, H. S., & Green, G. (2020). *Strategies and Tactics of Behavioral Research and Practice* (Fourth edition. ed., Vol. 1). Routledge. https://doi.org/10.4324/9781315537085

Kaufman, J. S., Jaser, S. S., Vaughan, E. L., Reynolds, J. S., Di Donato, J., Bernard, S. N., & Hernandez-Brereton, M. (2010). Patterns in Office Referral Data by Grade, Race/Ethnicity, and Gender. *Journal of positive behavior interventions*, *12*(1), 44-54. https://doi.org/10.1177/1098300708329710

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry*, *35*(2), 259-281.

Kirkpatrick, B. A., Wright, S., Daniels, S., Taylor, K. L., McCurdy, M., & Skinner, C. H. (2019). Tootling in an After-School Setting: Decreasing Antisocial Interactions in At-Risk Students. *Journal of positive behavior interventions*, *21*(4), 228-237. https://doi.org/10.1177/1098300719851226

Kohn, A. (2006). *Unconditional parenting: Moving from rewards and punishments to love and reason*. Simon and Schuster.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What works clearinghouse*.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-Case Intervention Research Design Standards. *Remedial and special education*, *34*(1), 26-38. https://doi.org/10.1177/0741932512452794

Kratochwill, T. R., Horner, R. H., Levin, J. R., Machalicek, W., Ferron, J., & Johnson, A. (2021). Single-case design standards: An update and proposed upgrades. *Journal of school psychology*, *89*, 91-105. https://doi.org/https://doi.org/10.1016/j.jsp.2021.10.006

Kvalsund, R., & Bele, I. V. (2010). Students with special educational needs—Social inclusion or marginalisation? Factors of risk and resilience in the transition between school and early adult life. *Scandinavian Journal of Educational Research*, *54*(1), 15-35.

Lambert, A. M. (2014). *Evaluating the Use of Tootling for Improving Upper Elementary/Middle School Students' Disruptive and Appropriate Behavior* [Doctoral Dissertation, The University of Southern Mississippi].

Lambert, A. M., Tingstrom, D. H., Sterling, H. E., Dufrene, B. A., & Lynne, S. (2015). Effects of Tootling on Classwide Disruptive and Appropriate Behavior of Upper-Elementary Students. *Behavior modification*, *39*(3), 413-430. https://doi.org/10.1177/0145445514566506

Lassen, S. R., Steele, M. M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to academic achievement in an urban middle school. *Psychology in the schools*, *43*(6), 701-712.

Lerman, D. C., Iwata, B. A., Shore, B. A., & DeLeon, I. G. (1997). Effects of intermittent punishment on self-injurious behavior: An evaluation of schedule thinning. *Journal of applied behavior analysis*, *30*(2), 187-201.

Lewis, T. J., Jones, S. E., Horner, R. H., & Sugai, G. (2010). School-wide positive behavior support and students with emotional/behavioral disorders: Implications for prevention, identification and intervention. *Exceptionality*, *18*(2), 82-93.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care

interventions: explanation and elaboration. *PLoS Med*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lillard, A. S. (2005). *Montessori: The science behind the genius*. Oxford Press.

Lipscomb, A. H., Anderson, M., & Gadke, D. L. (2018). Comparing the effects of ClassDojo with and without Tootling intervention in a postsecondary special education classroom setting. *Psychology in the schools*, *55*(10), 1287-1301. https://doi.org/10.1002/pits.22185

Litoe, L., & Pumroy, D. K. (1975). A brief review of classroom group-oriented contingencies. *Journal of applied behavior analysis*, *8*(3), 341.

Lum, J. D. K., Radley, K. C., Tingstrom, D. H., Dufrene, B. A., Olmi, D. J., & Wright, S. J. (2019). Tootling With a Randomized Independent Group Contingency to Improve High School Classwide Behavior. *Journal of positive behavior interventions*, *21*(2), 93-105. https://doi.org/10.1177/1098300718792663

Lum, J. D. K., Tingstrom, D. H., Dufrene, B. A., Radley, K. C., & Lynne, S. (2017). Effects of Tootling on Classwide Disruptive and Academically Engaged Behavior of General-Education High School Students. *Psychology in the schools*, *54*(4), 370-384. https://doi.org/10.1002/pits.22002

Lynne Lane, K., Stanton-Chapman, T., Roorbach Jamison, K., & Phillips, A. (2007). Teacher and Parent Expectations of Preschoolers' Behavior:Social Skills Necessary for Success. *Topics in Early Childhood Special Education*, *27*(2), 86-97. https://doi.org/10.1177/02711214070270020401

Madden, N. A., & Slavin, R. E. (1983). Effects of cooperative learning on the social acceptance of mainstreamed academically handicapped students. *The Journal of Special Education*, *17*(2), 171-182.

Mangiapanello, K. A., & Hemmes, N. S. (2015). An Analysis of Feedback from a Behavior Analytic Perspective. *The Behavior Analyst*, *38*(1), 51-75. https://doi.org/10.1007/s40614-014-0026-x

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional psychology: Research and practice*, *16*(2), 191.

McHugh, M. B., Tingstrom, D. H., Radley, K. C., Barry, C. T., & Walker, K. M. (2016). Effects of Tootling on Classwide and Individual Disruptive and Academically

Engaged Behavior of Lower-Elementary Students. *Behavioral interventions*, *31*(4), 332-354. https://doi.org/10.1002/bin.1447

Mhuru, M. (2020). *He Whakaaro: The educational experiences of disabled learners*. Education Counts

Miller, J. M. (2017). *Evaluating the Effectiveness of the Tootling Intervention in the Preschool Setting* [Master's Thesis, The University of Utah].

Montessori Jr, M. M. (1976). Education for Human Development: Understanding Montessori.

Morley, S., Masterson, C., & Main, C. J. (2017). Visual analysis of single-case data. In *Single Case Methods in Clinical Psychology* (pp. 87-125). Routledge.

Morris, J. K. (2024). *The Effect of Tootling on Academic Engagement and Prosocial Behavior* [Ed.S School Psychology, University of Nebraska at Omaha].

Morrison, J. Q., & Jones, K. M. (2007). The effects of positive peer reporting as a class-wide positive behavior support. *Journal of behavioral education*, *16*, 111-124.

Murphy, J., & Zlomke, K. (2014). Positive peer reporting in the classroom: A review of intervention procedures. *Behavior analysis in practice*, *7*, 126-137.

Närhi, V., Kiiski, T., & Savolainen, H. (2017). Reducing disruptive behaviours and improving classroom behavioural climate with class-wide positive behaviour support in middle schools. *British educational research journal*, *43*(6), 1186-1205.

New Zealand Ministry of Education. (2015). *Positive Behaviour for Learning: School-Wide Tier One Implementatio Manual*.

Ng, C., Bartlett, B., & Elliott, S. N. (2018). Empowering Engagement. *Creating Learning Opportunities for Students from Challenging Backgrounds*.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior therapy*, *40*(4), 357-367.

Powell, C. (2020). *Use your kind words please: Effects of tootling on student and teacher behaviour in New Zealand primary-school classrooms* [Master's Thesis, The University of Waikato].

Ray, J. E. (2019). *Evaluating the Effects of Tootling When Implemented in Special Education Classrooms Providing Behavior Supports* [Doctoral Dissertation, Minnesota State University, Mankato].

Reddy, L. A., Fabiano, G. A., Dudek, C. M., & Hsu, L. (2013). Instructional and behavior management practices implemented by elementary general education teachers.

Journal of school psychology, *51*(6), 683-700. https://doi.org/https://doi.org/10.1016/j.jsp.2013.10.001

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-Level Positive Behavior Supports in Schools Implementing SW-PBIS: Identifying Areas for Enhancement. *Journal of positive behavior interventions*, *15*(1), 39-50. https://doi.org/10.1177/1098300712459079

Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher use of behavior-specific praise. *Behavior modification*, *31*(3), 247-263.

Riley, J. L., McKevitt, B. C., Shriver, M. D., & Allen, K. D. (2011). Increasing On-Task Behavior Using Teacher Attention Delivered on a Fixed-Time Schedule. *Journal of behavioral education*, *20*(3), 149-162. https://doi.org/10.1007/s10864-011-9132-y

Robers, S., Kemp, J., Rathbun, A., & Morgan, R. E. (2014). Indicators of School Crime and Safety: 2013. NCES 2014-042/NCJ 243299. *National Center for Education Statistics*.

Ruiz-Olivares, R., Pino, M. J., & Herruzo, J. (2010). Reduction of disruptive behaviors using an intervention based on the Good Behavior Game and the Say-Do-Report Correspondence. *Psychology in the schools*, *47*(10), 1046-1058. https://doi.org/10.1002/pits.20523

Salinas, E. A. (2021). *Evaluating the Effects of Tootling on Disruptive and Academically Engaged Behaviors of Students with Emotional and Behavioral Disorders* [Doctoral Dissertation, The University of Texas at San Antonio].

Sapon-Shevin, M. (1994). Cooperative learning and middle schools: What would it take to really do it right? *Theory into practice*, *33*(3), 183-190.

Schakel, J. A., & Darveaux, D. X. (1984). The good behavior game plus merit: Controlling disruptive behavior and improving student motivation. *School psychology review*, *13*(4), 510-514.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and special education*, *8*(2), 24-33.

Segrin, C., & Flora, J. (2000). Poor social skills are a vulnerability factor in the development of psychosocial problems. *Human communication research*, *26*(3), 489-514.

Shernoff, E. S., Lekwa, A. L., Reddy, L. A., & Davis, W. (2020). Teachers' Use and Beliefs About Praise: A Mixed-Methods Study. *School psychology review*, *49*(3), 256-274. https://doi.org/10.1080/2372966X.2020.1732146

Sigafoos, J., Lang, R., Rispoli, M., & Matson, J. L. (2021). Definition and Rationale for Functional Assessment. In (pp. 3-23). Springer International Publishing. https://doi.org/10.1007/978-3-030-66270-7_1

Skinner, C. H. (2002). An empirical analysis of interspersal research: Evidence, implications, and applications of the discrete task completion hypothesis. *Journal of school psychology*, *40*(4), 347-368.

Skinner, C. H., Cashwell, T. H., & Skinner, A. L. (2000). Increasing tootling: The effects of a peer-monitored group contingency program on students' reports of peers' prosocial behaviors. *Psychology in the schools*, *37*(3), 263-270. https://doi.org/10.1002/(SICI)1520-6807(200005)37:3<263::AID-PITS6>3.0.CO;2-C

Skinner, C. H., Skinner, C. F., Skinner, A. L., & Cashwell, T. H. (1999). Using interdependent contingencies with groups of students: Why the principal kissed a pig. *Educational Administration Quarterly*, *35*(5), 806-820.

Slavin, R. E. (1991). Cooperative learning and group contingencies. *Journal of behavioral education*, *1*, 105-115.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994). 'Whenever and Wherever We Choose': The Replication of'Success for All'. *The Phi Delta Kappan*, *75*(8), 639-647.

Smith, S. M., Simon, J., & Bramlett, R. K. (2009). Effects of Positive Peer Reporting (PPR) on Social Acceptance and Negative Behaviors Among Peer-Rejected Preschool Children. *Journal of applied school psychology*, *25*(4), 323-341. https://doi.org/10.1080/15377900802595359

Stormont, M. A., Smith, S. C., & Lewis, T. J. (2007). Teacher implementation of precorrection and praise statements in Head Start classrooms as a component of a program-wide system of positive behavior support. *Journal of behavioral education*, *16*, 280-290.

Sugai, G., & Horner, R. (2002). The Evolution of Discipline Practices: School-Wide Positive Behavior Supports. *Child & family behavior therapy*, *24*(1-2), 23-50. https://doi.org/10.1300/J019v24n01_03

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-Intervention and School-Wide Positive Behavior Supports: Integration of Multi-Tiered System Approaches. *Exceptionality : the official journal of the Division for Research of the Council for Exceptional Children*, *17*(4), 223-237. https://doi.org/10.1080/09362830903235375

Suldo, S. M., Gormley, M. J., DuPaul, G. J., & Anderson-Butcher, D. (2014). The impact of school mental health on student and school-level academic outcomes: Current status of the research and future directions. *School Mental Health*, *6*, 84-98.

Sun, C. M. (2007). The Impact of Inclusion-Based Education on the Likelihood of Independence for Today's Students With Special Needs. *Journal of Special Education Leadership*, *20*(2).

Szumski, G., Smogorzewska, J., & Grygiel, P. (2022). Academic achievement of students without special educational needs and disabilities in inclusive education–Does the type of inclusion matter? *Plos one*, *17*(7), e0270124.

Tekin-Iftar, E., Olcay-Gul, S., & Collins, B. C. (2019). Descriptive Analysis and Meta Analysis of Studies Investigating the Effectiveness of Simultaneous Prompting Procedure. *Exceptional children*, *85*(3), 309-328. https://doi.org/10.1177/0014402918795702

Thoele, J. M., & Sayeski, K. L. (2024). The effects of student- and teacher-led tootling on student disruptive behavior. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1449306

Thomas, J. D., Presland, I. E., Grant, M. D., & Glynn, T. L. (1978). Natural Rates of Teacher Approval and Disapproval in Grade-7 Classrooms. *Journal of applied behavior analysis*, *11*(1), 91-94.

Towl, P. (2013). Making opportunity from disappointment: Students, parents and teachers talk about stand-down. *New Zealand journal of educational studies*, *48*(1), 127-139.

Tran, K. H. (2024). *A Comparison of Reward Criteria in Tootling* [Master's Thesis, California State University].

UNESCO. (2017). *A guide for ensuring inclusion and equity in education*. Paris: UNESCO

Valencia, R. R., & Suzuki, L. A. (2000). *Intelligence testing and minority students: Foundations, performance factors, and assessment issues* (Vol. 3). Sage.

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development*, *93*(4), 403-411. https://doi.org/10.1002/jcad.12038

Vollmer, T. R., Hagopian, L. P., Bailey, J. S., Dorsey, M. F., Hanley, G. P., Lennox, D., Riordan, M. M., & Spreat, S. (2011). The association for behavior analysis international position statement on restraint and seclusion. *Behav Anal*, *34*(1), 103-110. https://doi.org/10.1007/bf03392238

Von Brock, M. B., & Elliott, S. N. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of school psychology*, *25*(2), 131-144.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices*. Wadsworth.

Wheldall, K., & Merrett, F. (1988). Which Classroom Behaviours do Primary School Teachers say they find most Troublesome? *Educational Review*, *40*(1), 13-27. https://doi.org/10.1080/0013191880400102

White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of applied behavior analysis*, *8*(4), 367-372.

Whitefield, B. L., Harry, S. W., McIntosh, D. E., McIntosh, C. E., Chitiyo, A., English, A. M., & Poore, B. R. (2025). Elementary School Tootling: Implications on Notecard Responding and Teacher Observed Accuracy. *Journal of applied school psychology*, *41*(1), 44-71. https://doi.org/10.1080/15377903.2024.2429815

Wills, R. (2006). Special education 2000: A New Zealand experiment. *International journal of inclusive education*, *10*(02-03), 189-199.

Witt, J. C., & Elliott, S. N. (1985). *Acceptability of classroom intervention strategies*.

Witt, J. C., & Martens, B. (1983). Intervention rating profile. *School Psychology Quarterly*.

Wolfe, K., Barton, E. E., & Meadan, H. (2019). Systematic protocols for the visual analysis of single-case research data. *Behavior analysis in practice*, *12*, 491-502.

Wolfe, K., Seaman, M. A., & Drasgow, E. (2016). Interrater agreement on the visual analysis of individual tiers and functional relations in multiple baseline designs. *Behavior modification*, *40*(6), 852-873.

Wright, R. A., & McCurdy, B. L. (2012). Class-wide positive behavior support and group contingencies: Examining a positive variation of the good behavior game. *Journal of positive behavior interventions*, *14*(3), 173-180.

Wright, S., Skinner, C. H., Crewdson, M., Moore, T., McCurdy, M., Coleman, M. B., & Rocconi, L. M. (2022). Using A Tootling Intervention to Sequentially Enhance and

Maintain At-Risk Elementary Students' Performance of Multiple Social Skills. *School psychology*, *37*(3), 248-258. https://doi.org/10.1037/spq0000499

Wright, S., Skinner, C. H., Kirkpatrick, B. A., Daniels, S., Moore, T., & Crewdson, M. (2021). Using Tootling to Enhance First-Grade Students' Use of a Social Skill: Evaluating the Catching Compliments Game. *Education & treatment of children*, *44*(2), 101-113. https://doi.org/10.1007/s43494-021-00039-1

Wright, S. J. (2019). *Peer-Mediated Tootling with a Standardized Form and a Mystery Motivator in High School Classrooms* [Doctoral Dissertation, The University of Southern Mississippi].

Appendix A

Inter-Observer Agreement Form for Screening Out Duplicates.

Name of the observer: Jess Symes_____ Date of observation___23.1.25_____

#	Record title	Yes/No
1	Alstead, K. J. (2022). *Investigating the Effectiveness of the Positive Reinforcement Components of Tootling*. Minnesota State University, Mankato.	Yes
2	Chaffee, R. K., Briesch, A. M., Volpe, R. J., Johnson, A. H., & Dudley, L. (2020). Effects of a class-wide positive peer reporting intervention on middle school student behavior [Empirical Study; Quantitative Study]. *Behavioral Disorders*, *45*(4), 224-237.	Yes
3	Crewdson, M., Skinner, C. H., Wilson, C., Fowler, K., Richardson, R., & Wilhoit, B. (2023). Enhancing first-grade students' social skill performance: The interaction of Tootling and activities on social skill engagement [Empirical Study; Quantitative Study]. *Psychology in the Schools*, *60*(10), 4061-4078.	Yes
4	Crewdson, M. A., Richardson, R. D., Fowler, K., Skinner, C. H., Wright, S., & Cihak, D. (2024). Supplementing social skills training with Tootling to simultaneously enhance first-grade students' performance of two social skills [Empirical Study; Quantitative Study]. *School Psychology Review*, *53*(3), 236-250.	Yes
5	Enright, C. J. (2023). *An Investigation into the Effects of Tootling for Instruction-Following and On-Task Student Behaviour in a New Zealand Primary School* The University of Waikato].	Yes
6	Fowler, K. (2023). Altering Activities to Enhance the Effects of Social Skills Training and Tootling on Elementary Students Performance of Recently-Trained Social Skills.	Yes
7	Harry, S. W., Tingstrom, D. H., Dufrene, B. A., Dart, E. H., Radley, K. C., Lum, J. D., & Bates-Brantley, K. E. (2023). The effects of Tootling combined with public posting in high school classrooms [Empirical Study; Interview; Quantitative Study]. *Journal of Behavioral Education*, *32*(3), 565-584.	Yes
8	Hilt-Panahon, A., & Alstead, K. (2023). Components of Tootling. In *Implementing Tootling Interventions* (pp. 28-41). Routledge.	Yes
9	Hilt-Panahon, A., & Alstead, K. (2023). How to Implement Tootling. In *Implementing Tootling Interventions* (pp. 62-70). Routledge.	Yes
10	Kirkpatrick, B. A., Wright, S., Daniels, S., Taylor, K. L., McCurdy, M., & Skinner, C. H. (2019). Tootling in an after-school setting: Decreasing antisocial interactions in at-risk students [Empirical Study; Quantitative Study]. *Journal of Positive Behavior Interventions*, *21*(4), 228-237.	Yes
11	Lastrapes, R. E. (2016). Let Us Play: Using Research-Based Games to Facilitate Effective Instruction [Article]. *Beyond Behavior*, *25*(3), 27-33.	Yes

Continued on the next page

Inter-Observer Agreement Form for Screening Out Duplicates (cont.).

#	Record title	Yes/No
12	Lum, J. D., Radley, K. C., Tingstrom, D. H., Dufrene, B. A., Olmi, D., & Wright, S. J. (2019). Tootling with a randomized independent group contingency to improve high school classwide behavior [Empirical Study; Quantitative Study]. *Journal of Positive Behavior Interventions*, *21*(2), 93-105.	Yes
13	Lum, J. D. K. (2017). *Tootling with a randomized independent group contingency in a high school setting*. The University of Southern Mississippi.	Yes
14	McHugh, M. B., Tingstrom, D. H., Radley, K. C., Barry, C. T., & Walker, K. M. (2016). Effects of Tootling on classwide and individual disruptive and academically engaged behavior of lower-elementary students [Empirical Study; Quantitative Study]. *Behavioral Interventions*, *31*(4), 332-354.	Yes
15	Powell, C. (2020). *Use your kind words please: Effects of Tootling on student and teacher behaviour in New Zealand primary-school classrooms* The University of Waikato].	Yes
16	Shelton-Quinn, A. (2009). Increasing positive peer reporting and on-task behavior using a peer monitoring interdependent group contingency program with public posting [Dissertation Empirical Study; Quantitative Study]. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *70*(4-A), 1169.	Yes
17	Tran, K. H. (2024). *A COMPARISON OF REWARD CRITERIA IN TOOTLING.* California State University, Stanislaus].	Yes
18	Wright, S., Skinner, C. H., Kirkpatrick, B. A., Daniels, S., Moore, T., & Crewdson, M. (2021). Using Tootling to enhance first-grade students' use of a social skill: Evaluating the catching compliments game [Empirical Study; Quantitative Study]. *Education & Treatment of Children*, *44*(2), 101-113.	Yes

Observer's signature: 

Appendix B

Inter-Observer Agreement Form for Screening Out Non-Experimental Articles.

Name of the observer: Jess Symes_____ Date of observation___25.1.25_____

#	Record title	Yes/No
1	Chaffee, R. K. (2019). Effects of a class-wide positive peer reporting intervention on middle school student behavior [Dissertation Empirical Study; Quantitative Study]. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *80*(7-A(E)), No Pagination Specified.	Yes
2	Cihak, D., Kirk, E., & Boon, R. (2009). Effects of Classwide Positive Peer " Tootling" to Reduce the Disruptive Classroom Behaviors of Elementary Students with and without Disabilities [Article]. *Journal of Behavioral Education*, *18*(4), 267-278.	Yes
3	Derieux, J. (2019). Is the written component of Tootling effective? A comparison of the group contingency to a comparison writing procedure.	Yes
4	Enright, C. J. (2023). *An Investigation into the Effects of Tootling for Instruction-Following and On-Task Student Behaviour in a New Zealand Primary School* The University of Waikato].	Yes
5	Harry, S. W., Tingstrom, D. H., Dufrene, B. A., Dart, E. H., Radley, K. C., Lum, J. D., & Bates-Brantley, K. E. (2023). The effects of Tootling combined with public posting in high school classrooms [Empirical Study; Interview; Quantitative Study]. *Journal of Behavioral Education*, *32*(3), 565-584.	Yes
6	Lambert, A. M., Tingstrom, D. H., Sterling, H. E., Dufrene, B. A., & Lynne, S. (2015). Effects of Tootling on classwide disruptive and appropriate behavior of upper-elementary students [Empirical Study; Quantitative Study]. *Behavior Modification*, *39*(3), 413-430.	Yes
7	Lipscomb, A. H., Anderson, M., & Gadke, D. L. (2018). Comparing the effects of ClassDojo with and without Tootling intervention in a postsecondary special education classroom setting [Empirical Study; Quantitative Study]. *Psychology in the Schools*, *55*(10), 1287-1301.	Yes
8	Lum, J. D., Tingstrom, D. H., Dufrene, B. A., Radley, K. C., & Lynne, S. (2017). Effects of Tootling on classwide disruptive and academically engaged behavior of general-education high school students [Empirical Study; Followup Study; Interview; Quantitative Study]. *Psychology in the Schools*, *54*(4), 370-384.	Yes
9	McHugh, M. B. (2014). The effects of the Tootling intervention using daily reinforcement.	Yes
10	McHugh, M. B., Tingstrom, D. H., Radley, K. C., Barry, C. T., & Walker, K. M. (2016). Effects of Tootling on classwide and individual disruptive and academically engaged behavior of lower-elementary students [Empirical Study; Quantitative Study]. *Behavioral Interventions*, *31*(4), 332-354.	Yes

Continued on the next page

Inter-Observer Agreement Form Screening Out Non-Experimental Articles (cont.).

#	Record title	Yes/No
11	Ray, J. E. (2019). *Evaluating the Effects of Tootling When Implemented in Special Education Classrooms Providing Behavior Supports*. Minnesota State University, Mankato.	Yes
12	Vandermey, C. (2013). *The use of Tootling as a classwide positive behavior support*	Yes
13	Wright, S., Skinner, C. H., Kirkpatrick, B. A., Daniels, S., Moore, T., & Crewdson, M. (2021). Using Tootling to enhance first-grade students' use of a social skill: Evaluating the catching compliments game [Empirical Study; Quantitative Study]. *Education & Treatment of Children*, *44*(2), 101-113.	Yes

Observer's signature 

Appendix C

Inter-Observer Agreement Form for Excluding Theses and Dissertations That Were Published as Peer-Reviewed Articles.

Name of the observer: Jess Symes_____ Date of observation___26.1.25_____

#	Record title	Yes/No
1	Agosta, A. L. (2009). *The effects of a class-wide group contingency to increase Tootling*	Yes
2	Derieux, J. (2019). Is the written component of Tootling effective? A comparison of the group contingency to a comparison writing procedure.	Yes
3	Fowler, K. (2023). Altering Activities to Enhance the Effects of Social Skills Training and Tootling on Elementary Students Performance of Recently-Trained Social Skills.	Yes
4	Kirkpatrick, B. A., Wright, S., Daniels, S., Taylor, K. L., McCurdy, M., & Skinner, C. H. (2019). Tootling in an after-school setting: Decreasing antisocial interactions in at-risk students [Empirical Study; Quantitative Study]. *Journal of Positive Behavior Interventions*, *21*(4), 228-237.	Yes
5	Lambert, A. M. (2014). *Evaluating the use of Tootling for improving upper elementary/middle school students' disruptive and appropriate behavior*. The University of Southern Mississippi.	Yes
6	Lipscomb, A. H., Anderson, M., & Gadke, D. L. (2018). Comparing the effects of ClassDojo with and without Tootling intervention in a postsecondary special education classroom setting [Empirical Study; Quantitative Study]. *Psychology in the Schools*, *55*(10), 1287-1301.	Yes
7	McHugh Dillon, M. B., Radley, K. C., Tingstrom, D. H., Dart, E. H., & Barry, C. T. (2019). The effects of Tootling via ClassDojo on student behavior in elementary classrooms [Empirical Study; Quantitative Study]. *School Psychology Review*, *48*(1), 18-30.	Yes
8	Morris, J. K. (2024). *The Effect of Tootling on Academic Engagement and Prosocial Behavior* University of Nebraska at Omaha].	Yes
9	Sherman, J. C. (2012). *Positive peer reporting and positive peer reporting combined with Tootling: A comparison of interventions*. The University of Southern Mississippi.	Yes
10	Whitefield, B. L., Harry, S. W., McIntosh, D. E., McIntosh, C. E., Chitiyo, A., English, A. M., & Poore, B. R. (2024). Elementary School Tootling: Implications on Notecard Responding and Teacher Observed Accuracy. *Journal of Applied School Psychology*, 1-28.	Yes
11	Wright, S. J. (2019). *Peer-mediated Tootling with a Standardized Form and a Mystery Motivator in High School Classrooms* The University of Southern Mississippi].	Yes

Observer's signature 

Appendix D

Inter-Observer Agreement Form for Applying Final Eligibility Criteria.

Eligibility criteria:

1. The study was conducted in a general education primary or secondary classroom;
2. The participants included general education classrooms with students with diagnosed special needs or students, receiving individual support;
3. Tootling was their primary independent variable;
4. Dependent variables included class-wide or individual behaviours, such as disruptive behaviours, academic engagement, on-task, passive off-task behaviours, or pro-social behaviours.

Name of the observer: Jess Symes Date of observation 26.1.25

#	**Record title**	**Yes/No**
1	Choundira, R. (2022). *Tootling through a cultural lens: Effects of Tootling on student and teacher behaviours in an inclusive school in Aotearoa* The University of Waikato].	Yes
2	Cihak, D., Kirk, E., & Boon, R. (2009). Effects of Classwide Positive Peer " Tootling" to Reduce the Disruptive Classroom Behaviors of Elementary Students with and without Disabilities [Article]. *Journal of Behavioral Education*, *18*(4), 267-278.	Yes
3	Harry, S. W., Tingstrom, D. H., Dufrene, B. A., Dart, E. H., Radley, K. C., Lum, J. D., & Bates-Brantley, K. E. (2023). The effects of Tootling combined with public posting in high school classrooms [Empirical Study; Interview; Quantitative Study]. *Journal of Behavioral Education*, *32*(3), 565-584.	Yes
4	Lambert, A. M., Tingstrom, D. H., Sterling, H. E., Dufrene, B. A., & Lynne, S. (2015). Effects of Tootling on classwide disruptive and appropriate behavior of upper-elementary students [Empirical Study; Quantitative Study]. *Behavior Modification*, *39*(3), 413-430.	Yes
5	Powell, C. (2020). *Use your kind words please: Effects of Tootling on student and teacher behaviour in New Zealand primary-school classrooms* The University of Waikato].	Yes
6	Thoele, J. M., & Sayeski, K. L. (2024). The effects of student-and teacher-led Tootling on student disruptive behavior. Frontiers in Education.	Yes

Observer's signature 

Appendix E

Quality Evaluation of Studies by a Second Observer.

Name of the observer: Jess Symes Date of observation___26.1.25_(revisited 22/2)

#	Study	Level of analysis	Participants	Research design	Study meets WWC standards
1	Lambert (2014)	Classroom	Classroom A	Withdrawal	With reservation (BLs only have 4 points each)
2	McHugh et al. (2016)	Classroom	Classroom B	Withdrawal	Yes
		Classroom	Classroom C	Withdrawal	With reservation (BL1 had 4 data-points, BL2 only had 3 points)
3	Dillon et al. (2019)	Classroom	Classroom A	Withdrawal	Yes
		Classroom	Classroom C		Yes
4	Powell (2020)	Classroom	Group 1	MB	Yes
		Classroom	Group 2		
		Classroom	Group 3		
		Student	Nikau	Withdrawal	No (AB not ABAB = doesn't meet 4 phase min.).
5	Enright (2023)	Classroom	Classroom	Withdrawal	Yes
		Student	Alea (TS3)	Withdrawal	Yes
6	Morris (2024)	Classroom	n/a	Withdrawal	With reservations (BL2 = 3 data points).
7	Whitefield et al. (2025)	Classroom	Classroom A	MB/AT	With reservations – Classroom A only has 4 BL data-points
			Classroom B		
			Classroom C		

Observer's signature: 

Appendix F

Inter-Observer Agreement Form for Systematic Visual Analysis of Classroom Level Studies.

Name of the observer: Jess Symes Date of observation: 11/2/25

#	Study	Score
1	Cihak et al. (2009), DB	4.5
2	Lambert (2014), Classroom A, DB	4.5
3	Lambert et al. (2015), Classroom B, DB	4
4	Lum et al. (2017), Classroom C, DB	4
5	Wright (2019), Classroom B, DB	3.5
6	Powell (2020), Groups 1,2,3, DB	4.11
7	Harry et al. (2023), Classrooms A, B, C, DB	1.22
8	Morris (2024), DB	3
9	McHugh et al. (2016), Classroom B, AB	4.75
10	Lum et al. (2017), Classroom B, AB	3.5
11	Lum et al. (2019), Classroom A, AB	5
12	Wright (2019), Classroom A, AB	4.5
13	McHugh Dillon et al. (2019), Classroom C, AB	5
14	Chaffee et al. (2020), Classroom A, AB	4
15	Harry et al. (2023), Classrooms A, B, C, AB	1.44
16	Enright (2023), AB	3.75
17	Derieux (2019), Classrooms A, B, C, POT	2.67
18	Lum et al. (2019), Classroom A, POT	2.25

Observer's signature 

Appendix G

Inter-Observer Agreement Form for Systematic Visual Analysis of Student Level Studies.

Name of the observer: Jess Symes Date of observation: 11/2/25

#	Study	Score
1	Enright (2023), Alea, AB	4.75
2	Salinas (2021), Student A-2, DB	4.75
3	Salinas (2021), Student C-1, DB	4.75
4	Salinas (2021), Student B-1, AB	4.75

Observer's signature: 

Appendix H

Analysis Tables for Classroom Level Studies.

Table H1

Classrooms characteristics across classroom level studies.

#	Study	Classroom	Total students	SN students	Types of special needs/ supports	PBIS school	Grade/ year	Age of students	Classroom level	Location
1	Cihak et al. (2009)	n/a	19	4	SLD, ADHD	no	3	8, 9	elementary	US
2	Lambert (2014)	Classroom A	28	4	SLD	yes	6	11, 12	elementary	US
3	Lambert et al. (2015)	Classroom B	17	2	SLD	yes	4	9, 10	elementary	US
4	McHugh et al. (2016)	Classroom B	21	3	OHI	yes	2	7, 8	elementary	US
		Classroom C	23	1	OHI	yes	3	8, 9	elementary	US
5	Lum et al. (2017)	Classroom A	29	5	OHI, SLD	no	9, 10	14, 15, 16	high	US
		Classroom B	22	4	OHI, SLD, ASD	no	9 , 10, 11, 12	14, 15, 16, 17, 18	high	US
		Classroom C	26	3	OHI, SLD	no	10,11,12	15, 16, 17, 18	high	US
6	Derieux (2019)	Classroom A	23	8	ASD, SLD, EMD, OHI	no	9 , 10, 11, 12	14, 15, 16, 17, 18	high	US
		Classroom B	19	9	ASD, SLD, OHI	no	9 , 10, 11, 12	14, 15, 16, 17, 18	high	US
		Classroom C	18	3	SLD, OHI	no	9 , 10, 11, 12	14, 15, 16, 17, 18	high	US
7	Dillon et al. (2019)	Classroom A	35	3	OHI-ADHD, ASD	no	5	10, 11	elementary	US
		Classroom C	19	5	OHI-ADHD, SLD	no	5	10, 11	elementary	US
8	Lum et al. (2019)	Classroom A	22	3	SLD	no	10,11,12	15, 16, 17, 18	high	US
		Classroom C	26	4	SLD, OHI	no	12	17, 18	high	US
9	Wright (2019)	Classroom A	21	3	SLD	no	9, 10, 11	14, 15, 16, 17	high	US
		Classroom B	23	4	SLD, OHI	no	9, 10, 11	14, 15, 16, 17	high	US
		Classroom C	18	8	SLD, OHI, Visual Impairment.	no	9, 10, 11	14, 15, 16, 17	high	US
10	Chaffee et al. (2020)	Classroom A	17	1	ADHD	no	6	11, 12	middle	US
11	Powell (2020)	Group 2	21	1	At-risk of EBD	no	1	5, 6	elementary	NZ
		Group 3	25	1	n/a	no	5,6	10, 11	elementary	NZ
12	Enright (2023)	n/a	23	At least 1	ASD	yes	3	7	elementary	NZ
13	Harry et al. (2023)	Classroom A	16	3	SLD, OHI, AU	no	9,10	14, 15, 16	high	US
		Classroom C	17	3	SLD, OHI, AU	no	9,10	14, 15, 16	high	US
		Classroom D	20	5	Language/speech impairments, SLD	no	9	14, 15, 16	high	US
14	Morris (2024)	n/a	22	4	Pull-in supports for reading instruction	no	3	9, 10	elementary	US
15	Tran (2024)	n/a	30	9	English Learners, in Resource*, Individualized Education Plans (IEP)	no	9	14, 15	high	US
16	Whitefield (2024)	Classroom A	17	6	Language/speech impairments	no	3	8, 9	elementary	US
		Classroom B	13	2	ED, Language impairment	no	4	9, 10	elementary	US
		Classroom C	18	4	SLD, OHI	no	5	10, 11	elementary	US

* SLD - Specific Learning Disability, OHI/OHD - Other Health Impairment/Disability, ASD - Autism Spectrum Disorder, ADHD - Attention Deficit Hyperactivity Disorder, EBD - Emotional Behaviour Disorder, EMD - Emotional Disability, DCD - Developmental Cognitive Disability.

Table H2

Research methodologies across classroom level studies.

#	Study	Research design	Screening	Fading	Follow up	Recording procedure	Group behaviour assessment	Observation session	DV	Treatment integrity	Effect size
1	Cihak et al. (2009)	withdrawal	no	no	no	ER	group	school day	DB	yes	no
2	Lambert (2014)	withdrawal	yes	no	no	10-s MTS	observation of students by groups and rotation of those groups	20 mins	DB, AB	yes	no
3	Lambert et al. (2015)	withdrawal/ MB	yes	no	yes	10-s MTS	observation of students by groups and rotation of those groups	20 mins	DB, AB	yes	NAP
4	McHugh et al. (2016)	withdrawal/ MB	yes	no	no	10-s PIR	students observed in rotation	20 mins	DB, AEB	yes	NAP
5	Lum et al. (2017)	withdrawal	yes	no	yes	10-second MTS	individual-fixed method	20 mins	DB, AEB	yes	NAP, Tau-U
6	Derieux (2019)	MB/AT	yes	no	no	10-s MTS	individual-fixed method	20 mins	AEB, DB, POT	yes	Baseline-Corrected Tau
7	Dillon et al. (2019)	withdrawal	yes	no	no	10-s MTS	individual-fixed method	20 mins	DB, AEB	yes	Baseline corrected Tau
8	Lum et al. (2019)	ABAB withdrawal	yes	no	no	10-s MTS	individual-fixed method	20 mins	DB, AEB, POT	yes	Tau-U
9	Wright (2019)	withdrawal	yes	no	no	10-s MTS	individual-fixed method	20 mins	AEB, DB	yes	Tau-U
10	Chaffee et al. (2020)	withdrawal	yes	yes	no	15-s MTS, PIR	individual-fixed method	20 mins	DB, AEB	yes	NAP, Tau-U
11	Powell (2020)	MB	no	no	yes	15-s MTS	individual-fixed method	36-42 mins	DB, OTB	yes	Tau-U
12	Enright (2023)	withdrawal	no	yes	yes	ER, 20-s MTS	two students were selected randomly at the start of each observation session	35 mins	OTB	yes	Tau-U
13	Harry et al. (2023)	MB	yes	no	no	10-s MTS	individual-fixed method	20 mins	AEB, DB, POT	yes	Tau-U
14	Morris (2024)	withdrawal	yes	no	no	WIR, ER, DBR	group	20 mins	DB, AEB	yes	PND
15	Tran (2024)	AT	yes	no	no	DR, ER	group	28 mins average	DB	yes	no
16	Whitefield (2024)	MB/AT	no	no	yes	10-s MTS	individual-fixed method	20-minute	DB, AEB	yes	Tau-U

*ER - event recording, MTS – momentary time sampling, WIR – whole interval recording, PIR – partial interval recording, DBR – disruptive behaviour rating scales.

Table H3

Research teams and institutions across classroom level studies.

#	Study	Research team	University	Type of the paper
1	Cihak et al. (2009)	David Cihak, Emily Kirk, Richard Boon	University of Tennessee, University of Georgia	Peer-reviewed article
2	Lambert (2014)	Abigail M. Lambert, Daniel H. Tingstrom, Brad A. Dufrene, D. Joe Olmi, Keith Radley, Maureen Ryan	University of Southern Mississippi	PhD dissertation
3	Lambert et al. (2015)	Abigail M. Lambert, Daniel H. Tingstrom, Heather Sterling, Brad Dufrene, Shauna Lynn	University of Southern Mississippi, Live Oak Behavioural Psychology, Central Louisiana Resource Center on Developmental Disabilities	Peer-reviewed article
4	McHugh et al. (2016)	McHugh, M. B., Tingstrom, D. H., Radley, K. C., Barry, C. T., & Walker, K. M.	The University of Southern Mississippi, Washington State University	Peer-reviewed article
5	Lum et al. (2017)	John Lum, Daniel Tingstrom, Brad Dufrene, Keith Radley, Shauna Lynn	The University of Southern Mississippi	Peer-reviewed article
6	Derieux (2019)	James Derieux. Approved by: Dr. Evan Dart, Committee Chair Dr. Daniel Tingstrom Dr. Keith Radley	The University of Southern Mississippi	Master's thesis
7	Dillon et al. (2019)	McHugh Dillon, M. B., Radley, K. C., Tingstrom, D. H., Dart, E. H., & Barry, C. T.	Kennedy Krieger Institute, University of Utah, University of Southern Mississippi, Washington State University	Peer-reviewed article
8	Lum et al. (2019)	John Lum, Keith Radley, Daniel Tingstrom, Brad Dufrene, Joe Olmi, Sarah Wright	The University of Southern Mississippi	Peer-reviewed article
9	Wright (2019)	Sarah Joan Wright, Dr. Daniel H. Tingstrom, Dr. Brad A. Dufrene, Dr. Evan H. Dart, Dr. Keith C. Radley, Dr. D. Joe Olmi, Dr. Karen S. Coats	The University of Southern Mississippi	PhD dissertation
10	Chaffee et al. (2020)	Ruth Chaffee, Amy Briesch, Robert Volpe, Austin Johnson, Laura Dudley	Northeastern University, University of California	Peer-reviewed article
11	Powell (2020)	Cheniel Powell, Angelika Anderson	University of Waikato	Master's thesis
12	Enright (2023)	Candace J. Enright, Angelika Anderson	University of Waikato	Master's thesis
13	Harry et al. (2023)	Sarah Harry, Daniel Tingstrom, Brad Dufrene, Evan Dart, Keith Radley, John Lum, Kayla Bates-Brentley	Ball State University, University of Southern Mississippi, University of South Florida, University of Utah, Vancouver School Board, Mississippi State University	Peer-reviewed article
14	Morris (2024)	Jaedyn K. Morris Dr. Lisa Kelly-Vance Dr. Brian McKevitt Dr. Ferial Pearson	University of Nebraska	Ed.S field project
15	Tran (2024)	Kieu Hanh Tran, Dr. Katie Wiskow, Dr. Shrinidhi Subramanium, Dr. William Potter	California State University	Master's thesis
16	Whitefield (2024)	Breya L. Whitefield, Sarah W. Harry, David E. McIntosh, Connie E. McIntosh, Argnue Chitiyo, Abel M. English & Brianna R. Poore	Ball State University	Peer-reviewed article

Table H4

Variations of Tootling procedure across classroom level studies.

Author and year	Tootling implemented all the time or assigned days/periods	Voting for the name of the intervention	Collaboration with students choosing the reward	Use of technology	Reporting time	Reporting limits per student per day	Standardised format of reports	Intended days to achieve the goal	Goal, number of tootles	Goals kept consistent or increased	Goal known to students	Goal set or randomly selected	"Total tootles" goal	"Students wrote tootles" goal	"Students received tootles" goal	Reading five tootles aloud to students	Public posting of progress	Group contingency	Tootling implementer
Cihak et al. (2009)	school day	no	yes	no	school day	no	no	1-2 days	75	consistent	yes	set	yes	no	no	yes	yes	interdependent	teacher
Lambert (2014)	1 class period	no	yes	no	1 class period	2 reports	no	3-5 days	60 - 100	increased initially then kept consistent	yes	set	yes	no	no	yes	yes	interdependent	teacher
Lambert et al. (2015)	2 hr blocks	no	yes for A, no for B	no	2 hr block	2 for a, 1 for B	no	several days	65 initially	increased by teachers	yes	set	yes	no	no	yes	yes	interdependent	teacher
McHugh et al. (2016)	20-30-min period for A and B, 1hr for C	no	yes	no	20-30-min period for A and B, 1hr for C	no	no	1 day	30	consistent	yes	set	yes	no	no	yes	yes	interdependent	teacher
Lum et al. (2017)	95 min blocks	yes	yes	no	95 min blocks	no	no	3 days to 2 weeks	40-85 with more for more desirable rewards	varied according to the goal	yes	set	yes	no	no	yes	yes	interdependent	teacher
Derieux (2019)	90-min blocks once in 3 days	no	no	no	during transition times	2 reports	no	1 day	number of students on the day multiplied by 2	changed depending on the number of students present on the day	yes	set	yes	no	no	no	yes	interdependent	Teacher + researcher
Dillon et al. (2019)	at least three times per week. 20-min intervention session	no	yes	ClassDojo	approx. 3 mins for a class at the end of the teaching period	5 sec per student	yes	1-2 times a week	150	consistent	yes	set	yes	no	no	no, but displayed on the screen	yes	interdependent	teacher
Lum et al. (2019)	95 min blocks	yes	yes	no	95 min blocks	no	no	1 day	at least 3 tootles about different students and by 2 different students	consistent	yes	set	no	yes	yes	yes	no	randomised independent	teacher
Wright (2019)	90-min blocks	yes	yes	no	during transition times	2 reports	yes	1 day + Mystery motivator	NR	NR	yes	set but access to the reward random	yes	no	no	yes	no	interdependent	student interventionists

Continues on the next page

Table H4

Variations of Tootling procedure across classroom level studies (cont.).

Author and year	Tootling implemented all the time or assigned days/periods	Voting for the name of the intervention	Collaboration with students choosing the reward	Use of technology	Reporting time	Reporting limits per student per day	Standardised format of reports	Intended days to achieve the goal	Goal, number of tootles	Goals kept consistent or increased	Goal known to students	Goal set or randomly selected	"Total tootles" goal	"Students wrote tootles" goal	"Students received tootles" goal	Reading five tootles aloud to students	Public posting of progress	Group contingency	Tootling implementer
Chaffee et al. (2020)	42-49 min blocks	yes	yes	no	NR	no	no	NR	50	NR	yes	set	yes	no	no	yes	yes	interdependent	teacher
Powell (2020)	school day	no	yes	no	school day	no	no	NR	60-90 for A, 40-70 for B, 25 for C	increased for A and B, consistent for C	yes	set	yes	no	no	yes	yes	interdependent	teacher
Enright (2023)	every day, first half of the day	no	yes	no	50-60 minutes during lunchtime	3 reports	yes	1 day and for 3 days to receive a reward	1 to the number of students in class that day	randomised	no	randomly selected	no	no	yes	yes	yes, the stickers	interdependent	teacher
Harry et al. (2023)	school day	yes	yes	no	transition times	2 reports	no	3 days	NR	consistent	yes	set	yes	no	no	no, but individual public posting in B+C phase	yes	interdependent	teacher
Morris (2024)	school day 5 days a week	no	yes	ClassDojo	5 minutes per student at the end of the teaching period	no	no	NR	NR	changed as discussed by the teacher and students	yes	set	yes	no	no	no	yes	interdependent	teacher
Tran (2024)	Tootling on 3 of 4 days, sessions 21-30 mins	yes	yes	no	at the end of the session	1 initially, then no limit	no	1 day	varied according to condition	10 to 60, double the number of students in class	yes	randomly selected	yes	10 to 30, or the number of students present in class that day	10 to 30, or the number of students present in class that day	yes initially, then no	no	interdependent	teacher + researcher
Whitefield (2024)	20-minute session	no	NR	no	transition times	2 reports	2 conditions	1 day	NR	NR	yes	NR	yes	no	no	yes	yes	interdependent	teacher

Table H5

Social validity of Tootling across classroom level studies.

#	Study	Classroom	SV measure for teachers	SV by teachers score	Socially valid and acceptable by teachers	SV measure for students	SV by students mean score	Socially valid and acceptable by students	Was Tootling carried on after the study
1	Cihak et al. (2009)	n/a	IRP-15	86	yes	NR	NR	NR	NR
2	Lambert (2014)	Classroom A	IRP-15	68	yes	CIRP (target students only)	38	yes	NR
3	Lambert et al. (2015)	Classroom B		85	yes		NR	NR	yes
4	McHugh et al. (2016)	Classroom B Classroom C	IRP-15	90 74	yes yes	CIRP (target students only)	4.7 3	yes yes	NR NR
5	Lum et al. (2017)	Classroom A Classroom B Classroom C	BIRS	4.08 5.58 5.08	yes yes yes	NR	NR NR NR	NR NR NR	no no no
6	Derieux (2019)	Classroom A Classroom B Classroom C	URP- IR	3.87* 4.39* 4.21*	yes yes yes	CURP	not presented due to few students returning the form	NR NR NR	NR NR NR
7	Dillon et al. (2019)	Classroom A Classroom C	BIRS	5.3 and 5.1 4.3	yes yes	NR	NR NR	NR NR	NR NR
8	Lum et al. (2019)	Classroom A Classroom C	BIRS	4.57 4.17	yes yes	CIRP	4.93 5.33	yes yes	NR NR
9	Wright (2019)	Classroom A Classroom B Classroom C	BIRS	4.04 4.84	yes yes yes	CIRP	5.46/32.75 4.48/26.89 4.37/26.20	yes yes yes	no no no
10	Chaffee et al. (2020)	Classroom A	URP-IR	4.55*	yes	CURP	3.32*	yes	NR
11	Powell (2020)	Group 2 Group 3	BIRS BIRS	4.96 4.74	yes yes	CIRP CIRP	35.16 31.91	yes yes	yes no
12	Enright (2023)	n/a	BIRS	3.97	yes	CIRP	4.8/35	yes	no
13	Harry et al. (2023)	Classroom A Classroom C Classroom D	BIRS	5.25 4.88&5.64 4.71	yes yes yes	CIRP	4.73/28.40 4.86/29.17 4.88/29.29	yes yes yes	NR NR NR
14	Morris (2024)	n/a	NR	NR	NR	NR	NR	NR	NR
15	Tran (2024)	n/a	URP-IR	4.69*	yes	CURP	3.4*	yes	NR
16	Whitefield (2024)	Classroom A Classroom B Classroom C	URP-I URP-I URP-I	NR NR NR	yes yes yes	CURP CURP CURP	NR NR NR	yes yes yes	NR NR NR

Table H6

The Presence and the size of the reported effect across classroom level studies.

#	Study	Classroom	Effect size value average, Tau-U if not stated			Presence and size of the effect		
			DB	AB	POT	DB	AB	POT
1	Cihak et al. (2009)	n/a	NR			yes		
2	Lambert (2014)	Classroom A	NR	NR		yes	yes	
3	Lambert et al. (2015)	Classroom B	0.94	0.95		large	large	
4	McHugh et al. (2016)	Classroom B	NAP 1	NAP 1		very large	very large	
		Classroom C	NAP 1	NAP 1		very large	very large	
5	Lum et al. (2017)	Classroom A	0.85	0.73		moderate	moderate	
		Classroom B	0.92	0.64		large	moderate	
		Classroom C	0.78	0.53		large	moderate	
6	Derieux (2019)	Classroom A	BC Tau 0.35	BC Tau 0.6	BC Tau 0.47	moderate	large	moderate
		Classroom B	BC Tau 0.11	BC Tau 0.43	BC Tau 0.56	small	moderate	moderate
		Classroom C	BC Tau 0.06	BC Tau 0.17	BC Tau 0.07	small	small	small
7	Dillon et al. (2019)	Classroom A	0.73	0.722		large	large	
		Classroom C	0.76	0.753		large	large	
8	Lum et al. (2019)	Classroom A	0.94	0.96	0.29	very large	very large	moderate
		Classroom C	0.87	0.93	0.13	very large	very large	small
9	Wright (2019)	Classroom A	0.31	0.88		moderate	very large	
		Classroom B	0.25	0.48		moderate	moderate	
		Classroom C	0.53	0.77		moderate	large	
10	Chaffee et al. (2020)	Classroom A	0.48	0.68		moderate	large	
11	Powell (2020)	Group 2	0.86	0.96		very large	very large	
		Group 3	0.05	0.84		small	very large	
12	Enright (2023)	n/a	PEM 100%	PEM 100%		yes	yes	
13	Harry et al. (2023)	Classroom A	1	1	0.36	very large	very large	moderate
		Classroom C	0.73	1	0.8	large	very large	large
		Classroom D	0.85	0.88	0.58	very large	very large	moderate
14	Morris (2024)	n/a		0.67			large	
15	Tran (2024)	n/a	PND 100%	PND 100%		yes	yes	
16	Whitefield (2024)	Classroom A	0.63	0.87		large	very large	
		Classroom B	0.67	0.95		large	very large	
		Classroom C	0.56	0.8		moderate	very large	

Table H7

The results of systematic visual analysis and comparison with effects reported in studies across classroom level studies.

#	Author and year	Classroom	Systematic visual analysis score			Effect present			Effect size value average, Tau-U if not stated			Interpretation of effect size		
			DB	AB	POT	DB	AB	POT	DB	AB	POT	DB	AB	POT
1	Cihak et al. (2009)	n/a	4.75			yes			NR			NR		
2	Lambert (2014)	Classroom A	4.5	4		yes	yes		NR	NR		NR	NR	
3	Lambert et al. (2015)	Classroom B	4.75	4.5		yes	yes		0.94	0.95		large	large	
4	McHugh et al. (2016)	Classroom B	4.5	5		yes	yes		NAP 1	NAP 1		very large	very large	
		Classroom C	5	5		yes	yes		NAP 1	NAP 1		very large	very large	
5	Lum et al. (2017)	Classroom A	4.25	3.75		yes	yes		0.85	0.73		moderate	moderate	
		Classroom B	4.25	4		yes	yes		0.92	0.64		large	moderate	
		Classroom C	4.75	3.75		yes	yes		0.78	0.53		large	moderate	
6	Derieux (2019)	Classrooms A, B, C	2.67	1.44	2.67	no	no	no	BC Tau 0.17	BC Tau 0.40	BC Tau 0.42	small	moderate	moderate
7	Dillon et al. (2019)	Classroom A	5	4.75		yes	yes		0.73	0.722		large	large	
		Classroom C	5	5		yes	yes		0.76	0.753		large	large	
8	Lum et al. (2019)	Classroom A	5	5	1.5	yes	yes	no	0.94	0.96	0.29	very large	very large	moderate
		Classroom C	4.75	4.75	2.25	yes	yes	no	0.87	0.93	0.13	very large	very large	small
9	Wright (2019)	Classroom A	4	4.5		yes	yes		0.31	0.88		moderate	very large	
		Classroom B	3.25	4		yes	yes		0.25	0.48		moderate	moderate	
		Classroom C	4	4		yes	yes		0.53	0.77		moderate	large	
10	Chaffee et al. (2020)	Classroom A	3.5	4		yes	yes		0.48	0.68		moderate	large	
11	Powell (2020)	Groups 1,2,3	2.89	4.78		no	yes		0.37	0.93		moderate	very large	

Continues on the next page

Table H7

The results of systematic visual analysis and comparison with effects reported in studies across classroom level data studies (cont.).

#	Author and year	Classroom	Systematic visual analysis score			Effect present			Effect size value average, Tau-U if not stated			Interpretation of effect size		
			DB	AB	POT	DB	AB	POT	DB	AB	POT	DB	AB	POT
12	Enright (2023)	n/a		3.5		no	yes			0.67			large	
13	Harry et al. (2023)	Classrooms A,B,C	1.48	1.65	1.11	no	no	no	0.82	0.87	0.62	very large	very large	large
14	Morris (2024)	n/a	2.75	4.25		no	yes		PND 100%	PND 100%		large	large	
16	Whitefield (2024)	Classrooms A,B,C	3.33	4.56		yes	yes		0.62	0.87		moderate	very large	

Appendix I

Analysis Tables for Student Level Studies.

Table I1

Student characteristics across student level studies.

#	Study	Students	Age of students	Grade/ year	Student's gender	Types of special needs/supports	Classroom level	Classroom type	SWPBIS school	Country
12	Enright (2023)	Alea (TS3)	7	NR	female	ASD	primary	general	yes	NZ
17	Salinas (2021)	Student A-1	10	4	male	ED disability	elementary	special	no	US
		Student A-2	9	4	male	ED disability	elementary	special	no	US
		Student B-1	13	7	male	ED disability	middle	special	no	US
		Student B-2	11	6	male	ED disability	middle	special	no	US
		Student C-1	16	9	male	ED disability	high	special	no	US
		Student C-2	17	10	male	ED disability	high	special	no	US
18	Thoele (2024)	Elian	10	4	male	ADHD	elementary	general	yes	US
		Asher	10	4	male	OHI - ADHD, SLD	elementary	general	yes	US
		Darnell	11	5	male	no documented disabilities but tier 2 support	elementary	general	yes	US
		Jonah	11	5	male	SLD	elementary	general	yes	US
		Xavier	10	5	male	no documented disabilities but referred to tier 2 support	elementary	general	yes	US

*ED disability – emotional disturbance disability

Table I2

Methodologies across student level studies.

#	Study	Research design	Screening for inclusion	Fading	Follow up	Recording procedure	Observation session	Dependent variables	Treatment integrity measured	Effect size measure
12	Enright (2023)	ABABCD withdrawal	no	yes	yes	20-s MTS	35 mins	OTB	yes	Tau-U
17	Salinas (2021)	ABAB withdrawal	yes	no	no	10-s PIR; 10-s WIR	20 mins	DB, AEB	yes	Tau-U, SMD
18	Thoele (2024)	ATD with a maintenance phase	yes	yes	no	10-s MTS	20 mins	DB, AEB, POT	yes	no

Table I3

Research teams across student level studies.

#	Study	Research team	Affiliated institution	Type of the paper
12	Enright (2023)	Candace J. Enright, Angelika Anderson	University of Waikato	Master's thesis
17	Salinas (2021)	Ernesto Alberto Salinas, Richard Boon, Iliana Alanis, Kathy Ewoldt, Leslie Neely	University of Texas	PhD dissertation
18	Thoele (2024)	Jillian M. Thoele, Kristin L. Sayeski	University of Georgia	Peer-reviewed article

Table I4

Variations of the Tootling procedure across student level studies.

#	Author and year	Classroom	Tootling implemented all the time or assigned days/periods	Voting for the name of the intervention	Collaboration with students choosing the reward	Use of technology	Reporting time	Reporting limits per student per day	Standardised format of reports	Intended days to achieve the goal	Goal, number of tootles	Goals kept consistent or increased
12	Enright (2023)	Alea (TS3)	every day, first half of the day	no	yes	no	50-60 minutes during lunchtime	3 reports	yes	1 day and for 3 days to receive a reward	1 to the number of students in class that day	randomised
17	Salinas (2021)	Student A-1 Student A-2 Student B-1 Student B-2 Student C-1 Student C-2	school day	no	yes, Google form survey	Padlet, Google sheet	any time during school day	no	no	NR	1st goal number of students*3	increased
18	Thoele (2024)	Elian Asher Darnell Jonah Xavier	during one class	no	yes	no	at the end of the class	no, but every student had a "shout-out" partner	yes yes no no no	twice a week	40 40 42 42 42	consistent

Table I4 (cont.)

#	Author and year	Classroom	Goal known to students	Goal set or randomly selected	"Total tootles" goal	"Students wrote tootles" goal	"Students received tootles" goal	Reading five tootles aloud to students	Public posting of progress	Group contingency	Who implemented Tootling
12	Enright (2023)	Alea (TS3)	no	randomly selected	no	no	yes	yes	yes, the stickers	interdependent	teacher
17	Salinas (2021)	Student A-1 Student A-2 Student B-1 Student B-2 Student C-1 Student C-2	yes	set	yes	no	no	no, but all tootles were displayed on the screen	yes, bar chart on the screen	interdependent	teacher
18	Thoele (2024)	Elian Asher Darnell Jonah Xavier	yes	set	yes	no	no	yes	yes. Also When classes met their goals, the teacher placed the Tootling slips in students' folders to go home to their parents.	interdependent	teacher-led and student-led conditions

Table I5

Reported effects of Tootling across student level studies.

#	Study	Students	Effect size value average, Tau-U if not stated			Size/presence of the effect		
			DB	AB	POT	DB	AB	POT
12	Enright (2023)	Alea (TS3)		0.96		very large	very large	
17	Salinas (2021)	Student A-1	1	1		very large	very large	
		Student A-2	1	1		very large	very large	
		Student B-1	1	1		very large	very large	
		Student B-2	1	1		very large	very large	
		Student C-1	1	1		very large	very large	
		Student C-2	0.95	1		very large	very large	
18	Thoele (2024)	Elian	NR	NR	NR	yes	yes	no
		Asher	NR	NR	NR	yes	yes	no
		Darnell	NR	NR	NR	yes	yes	no
		Jonah	NR	NR	NR	yes	yes	no
		Xavier	NR	NR	NR	yes	yes	no

Table I6

Social validity of Tootling across student level studies.

#	**Author and year**	**Students**			**Acceptability by teachers, value**	**Acceptability by teachers, interpretation**	**Acceptability by students value**	**Acceptability by students interpretation**	**Was Tootling carried on after the study**
12	Enright (2023)	Alea (TS3)	BIRS	CIRP	4	socially valid and acceptable	4.8 and 35*	socially valid and acceptable	no
17	Salinas (2021)	Student A-1	BIRS	Student Feedback of Tootling form	5	socially valid and acceptable	no score	answers showed mostly positive attitude towards Tootling	NR
		Student A-2							
		Student B-1			5.04	socially valid and acceptable			
		Student B-2							
		Student C-1			4.96	socially valid and acceptable			
		Student C-2					NR	NR	
18	Thoele (2024)	Elian	Tootling Efficacy Questionnaire, Treatment Acceptability Survey for Teachers	Treatment Acceptability Questionnaire for Students	5	socially valid and acceptable	no score	socially valid and acceptable	yes
		Asher						socially valid and acceptable	
		Darnell			5	socially valid and acceptable		socially valid and acceptable	yes
		Jonah						socially valid and acceptable	
		Xavier			4.25	socially valid and acceptable		socially valid and acceptable	yes

* - classroom scores

Table I7

Systematic visual analysis scores and effect sizes across classroom level studies.

#	Study	Students	Systematic visual analysis score		Effect size value average, Tau-U if not stated		Interpretation of effect size	
			DB	**AB**	**DB**	**AB**	**DB**	**AB**
12	Enright (2023)	Alea (TS3)		5		0.96		very large
17	Salinas (2021)	Student A-1	5	5	1	1	very large	very large
		Student A-2	5	5	1	1	very large	very large
		Student B-1	5	5	1	1	very large	very large
		Student B-2	5	5	1	1	very large	very large
		Student C-1	5	5	1	1	very large	very large
		Student C-2	5	5	0.95	1	very large	very large