Evidence-Based Interventions for Primary-aged Children exhibiting Externalising Behaviours: A Systematic Literature Review

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Science in Psychology

at

The University of Waikato

by

GINA LOUISE POWELL

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2025

Abstract

Background: Externalising disorders (EDs) are among the most common behavioural problems in childhood, with disruptive behaviours in Aotearoa New Zealand schools recently identified as a major concern. Despite widespread implementation of interventions, uncertainty remains about their effectiveness and alignment with developmental theory.

Objectives: This systematic review aimed to evaluate the effectiveness of interventions for primary-aged children (5–14 years) exhibiting EDs, assess the certainty of evidence, examine moderating factors of intervention outcomes, and consider implications for practice in the Aotearoa context.

Methods: Following PRISMA 2020 guidelines, eight databases and trial registries were searched. Eligible studies were randomised controlled trials (RCTs) of interventions targeting EDs in primary-aged children. Risk of bias was assessed using the Cochrane RoB 2.0 tool, and findings were synthesised narratively given heterogeneity across interventions and outcomes.

Results: Twenty-three RCTs met inclusion criteria, yielding 148 outcome measures. The majority (77.7%) were at high risk of bias, largely due to reliance on unblinded raters, insufficient reporting, and baseline imbalances. Sensitivity analyses excluding very low-certainty studies left only 17 outcomes, of which eight were rated as high certainty. Behavioural outcomes were heavily over-represented (67.5%), with relational and environmental outcomes underexplored. Moderator analyses indicated that interventions targeting single subgroups produced more favourable outcomes than multisystem approaches, though overall evidence quality limited confidence in these findings.

Conclusions: Current RCT evidence for ED interventions in primary-aged children is of low quality, restricting firm conclusions about effectiveness. The imbalance in outcome domains and mechanisms suggests an overemphasis on behaviourist approaches at the expense of relational and ecological factors. High-quality, large-scale RCTs are urgently needed, particularly those addressing classroom environments, relational mechanisms, and culturally grounded approaches relevant to Aotearoa.

Registration: The protocol for this review was registered with the Open Science Framework (OSF). https://osf.io/9afzx/

Preface and Acknowledgements

To my incredibly supportive husband, I could not have done this without you. You took on more than your fair share of life during this time and I thank you from the bottom of my heart. To my wonderful kids, Thea and Oliver, you have taught me so much about life and given me some deep insights which helped me see intervention effectiveness in a different light. You are both truly wonderful. And to my parents Greg and Raechel, I could not ask for a more rock solid foundation in my life. Thank you for always being there supporting me.

Table of Contents

Lists of Tables, Figures

Tables

Figures

Evidence-Based Interventions for Primary-Aged Children Exhibiting Externalising Behaviours: A Systematic Literature Review

Current Knowledge

The Education Review Office (ERO) of Aotearoa released a report in early 2024 highlighting an increase in challenging behaviours in schools, low teacher capability to manage them, and limited access to expert support (Education Review Office, 2024). The behaviours outlined in the ERO report were classified in this review as externalising disorders (EDs). This classification was informed by their alignment with the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) and the Hierarchical Taxonomy of Psychopathology (HiTOP). Developmental theories of EDs were then used to contextualise how outcomes were defined and categorised for the purpose of evaluating interventions.

Findings from the Education Review Office Report

Classroom behaviour in Aotearoa consistently ranks among the worst in the OECD (Chamberlain & Medina, 2020; Education Review Office, 2024, p. 2, 17–20, 32 & 119). Increasing levels of challenging behaviour were reported by 74% of teachers and principals, and these behaviours are negatively impacting staff and students (Bradshaw et al., 2010; Education Review Office, 2024, p. 3 & 15).

Teachers reported feeling insufficiently supported, inadequately trained, and struggling to find time to address disruptive behaviours. New teachers report having inadequate behaviour management training; thus, learning to handle students' behaviour occurs on the job (Education Review Office, 2024, p. 53). Managing behaviour was reported as becoming increasingly complex due to a rising demand to cater to a variety of individual needs associated with developmental disorders. This shift has led to the perception that the profession has moved its focus from teaching to managing behaviour. Furthermore, challenging behaviour was identified as a key reason teachers consider leaving the profession (Education Review Office, 2024, p. 51, 58).

With growing behavioural issues and inadequate behaviour management training for teachers, there ought to be sufficient access to external expert support. Unfortunately, this is not the case. A majority of teachers and principals across Aotearoa agree that timely expert support is essential; however, 39-49% find it difficult to access this support (Education Review Office, 2024, p. 4). The shortage of psychologists in Aotearoa is well-

documented (Burns & Reporter, 2024; Every-Palmer et al., 2024; Health New Zealand, Te Whatu Ora, 2024), with a growing number of children unable to access the professional help they need (Manatū Hauora, Ministry of Health, 2024). ERO stated in their report that increasing access to experts would ultimately require an increase in the supply of experts (Education Review Office, 2024, p. 123). However, this is a systemic solution that will take some time.

Operationally Defining Challenging Behaviours

In conducting an evidence-based review of intervention efficacy, behaviours reported by ERO must be systematically classified and defined to enable replicability of the identification of relevant interventions. The behaviours outlined in ERO's report were classified as EDs, by aligning the reported behaviours with like symptoms in HiTOP. Developmental theories of EDs then offer a contextual understanding of how EDs were operationally defined in this review to select appropriate outcome measures in which to assess the effectiveness of ED interventions. Prominent challenging behaviours identified in the ERO report, listed in order of prevalence, included "talking inappropriately in class…distracting others…refusing to follow instructions…damaging or taking property…[and] physically harming others" (Education Review Office, 2024, p. 20). These behaviours are associated with the symptoms and traits of EDs.

In the DSM-5, disorders considered to include disruptive behaviour problems (DBPs) are categorised across neurodevelopmental and externalising sections of the manual, which risks introducing selective inclusion of disorders deemed appropriate across different researchers (American Psychiatric Association, 2022). To avoid subjectivity, this review uses the HiTOP framework, which categorises disorders by symptomatology, thereby reducing bias in classifying the behaviours described in ERO's report. Alternatively, the HiTOP framework is an evidence-based framework that categorises disorders by symptomatology, making it a practical framework for reducing bias through its capacity to classify described behaviours into relevant clinical disorders (Kotov et al., 2017; Ringwald et al., 2023).

In the HiTOP framework, specific behaviours reported by ERO are listed as symptom components and maladaptive traits within the disinhibiting and antagonistic externalising spectra, which relate to the antisocial behaviour subfactor. Certain diagnosable disorders in children that belong to these categories include conduct disorder (CD), oppositional defiant

disorder (ODD), attention deficit hyperactivity disorder (ADHD), and intermittent explosive disorder (IED) (Kotov et al., 2017).

Disruptive and distracting behaviours, such as inappropriately talking in class, are comparable to symptoms of ADHD. Specifically, these include impulse control issues, hyperactivity, and difficulty in controlling attentional processes or maintaining attention. Behaviours involving physical harm and property damage are related to symptoms of CD, IED, and traits of ADHD. CD symptoms include deliberately destroying property and exhibiting aggression towards others. Likewise, IED symptoms involve emotional outbursts that lead to unintentional destruction and harm. Physical harm and destruction can also stem from symptoms associated with ADHD, namely hyperactivity and poor impulse control, more colloquially known as acting before thinking. Lastly, refusing to follow instructions can be linked to several EDs, but most significantly, it is associated with symptoms of ODD, which include being argumentative and defiant (American Psychiatric Association, 2022).

Furthermore, EDs can also be referred to as DBPs or emotional and behavioural disorders (EBDs) (Ogundele, 2018). Beyond specific disorders, studies referring to DBPs and EBDs were included in the search strategy to capture as many relevant studies as possible.

Externalising Disorders in Aotearoa

In Aotearoa, childhood EDs range in prevalence from 5% to 16.7% (Heywood & Fergusson, 2016; Manatū Hauora, Ministry of Health, 2024; Moffitt et al., 2015). As recently as 2023/24, about 1 in 9 children in Aotearoa aged 2 to 14 were reported to have behavioural and emotional problems, with a significantly higher rate, closer to 50% among disabled children. Alongside this growth in prevalence, the percentage of children unable to access the professional support they need has sadly increased from 4.8% in 2016/17 to 6.5% in 2023/24 (Manatū Hauora, Ministry of Health, 2024). This aligns with 70-74% of principals' and teachers' perceptions of worsening behaviour (Education Review Office, 2024, p. 33).

Impact

Challenging behaviours were viewed as having an adverse effect on both students and teachers in the ERO report. As a result, all students experienced less learning, educational progress, and enjoyment at school. Students exhibiting challenging behaviours were also at greater risk of poorer life outcomes (Education Review Office, 2024, p. 88).

Research supports this, showing that children with disruptive or antisocial behaviours, regardless of gender, are more likely to face adverse life outcomes later, including involvement in crime, substance abuse, and unemployment (Bradshaw et al., 2010; Fergusson et al., 2005). Over a quarter of teachers linked reduced physical and mental well-being to managing challenging behaviours, and 60% of teachers, along with 82% of principals, reported significant increases in stress due to behaviour management. Unsurprisingly, over half of teachers stated that behaviour issues in schools greatly affect their enjoyment of the job and their desire to stay in the profession (Education Review Office, 2024, pp. 63–64). Similar effects are seen in families of children with EDs, where marital relationships often suffer, and family conflict, feelings of being overwhelmed, and helplessness tend to rise (Burt et al., 2003; Sajadi et al., 2020; Samek & Hicks, 2014). The long-term consequences of untreated EDs extend beyond students, teachers, and families to social and economic impacts in the wider community, with disproportionately high utilisation of health, justice, and welfare services. For example, 50% of those with CD are affected compared to 11.3% of the general population (Rivenbark et al., 2018).

Developmental Theories

The field of developmental psychopathology involves recognising a wide range of risk and protective factors (Venta et al., 2021). These include predictive or associated biopsychosocial factors such as genetics, neurobiological markers, environmental influences, interpersonal relationships, cognitive processes, and temperament, which have been found to either elevate or mitigate a child's risk of developing disorders (Eriksson et al., 2011; Holmes et al., 2001). Within the framework of developmental theories like Bandura's (1986) Social Cognitive Theory, Sameroff's (2009) Transactional Model of Development, and Bronfenbrenner's (2000) Ecological Systems Theory, it is clear that EDs do not develop in isolation—solely within the child—rather, a multitude of factors spanning biological and environmental domains collectively contribute (Sameroff, 2010). Children with EDs are not solely responsible for their condition. A thorough understanding of these influential factors, combined with developmental theories, is essential for making informed decisions regarding intervention strategies for EDs, including the validation of intervention targets and mechanisms of change.

Biopsychosocial Risk and Protective Factors. In the broader environmental context, prosocial peers, a well-organised neighbourhood, and adequate shelter are established protective factors (Eriksson et al., 2011). Conversely, poverty and poor social networks are identified as risk factors contributing to the development of EDs (Fergusson et al., 2005).

In the context of home, confirmed risk factors for EDs include exposure to violence and maltreatment through unpredictable, inconsistent, harsh, and abusive parenting practices (Braga et al., 2018; Fergusson et al., 2005). Conversely, secure attachment, a stable familial environment, and engaged, interested parents who exhibit predictable and affectionate parenting practices serve as protective factors (Buchanan & Flouri, 2001; Navarro-Soria et al., 2020). Notably, parenting styles are categorised into four types: authoritative (firm but fair), authoritarian (firm), permissive (indulgent), and neglectful (indifferent/absent) (Baumrind & Black, 1967). Of these, authoritative parenting is most associated with more positive outcomes (Lamborn et al., 1991). It balances monitoring child behaviour with being involved and interested, and it balances consistency in holding clear boundaries with being open to reasoning and supporting child psychological autonomy (Gray & Steinberg, 1999).

A child's temperament, which has demonstrated some stability throughout development, is a confirmed risk factor for EDs (Caspi et al., 1995). Children who develop EDs often display traits such as reactivity, aggression, impulsivity, or inattention. These traits influence their interpersonal relationships. A child's easy-going or difficult disposition affects how parents interact with them (Maccoby, 1984).

Within the brain and cognition, there are a range of influential factors in the development of EDs. Genetics can buffer maltreatment, with higher levels of Monoamine oxidase A in children showing some evidence of being a protective factor against the effects of maltreatment (Caspi et al., 2002). Neurological factors such as the volume of white and grey matter, lower serotonin levels in the brain, and structural differences affecting neural processes that regulate emotions, attention, and responses are all associated with EDs (Blair et al., 2021; Klasen et al., 2019; Olivier, 2004). Cognitive skills, including effortful control, executive functioning, healthy working memory, attentional systems, IQ, and language development, are established protective factors (Atherton et al., 2020; Holmes et al., 2001).

Given the variety of biopsychosocial influential factors, interventions should clearly specify which factors are targeted and how – especially when interventions focus on a child's behaviour, as the child may have multiple risk factors beyond their own behaviour working against them.

Social Cognitive Theory. Simply observing others' behaviours can lead to imitation, as supported by the Bobo Doll Experiment, where children mimicked the violent or kind behaviours adults displayed toward the Bobo doll (Bandura et al., 1961). Behaviours can also be reinforced through reciprocal determinism – a triadic interaction between a person's behaviours, cognition, and environment. Tsomokos & Flouri (2024) found that superior social cognitive abilities predicted lower emotional problems, such as understanding that others may have different perspectives on an interaction, especially in strained social enhancement situations. Improving social cognition and skills improves adaptive behaviours (Gresham & Elliott, 1987). Conversely, poor social cognition and skills are linked to the development and presence of EDs (Baumel et al., 2021; Dall et al., 2022; Sharp et al., 2011; Taubner et al., 2012). Those with ED are more likely to interpret others' intentions as hostile and, therefore, more prone to engage in conflict (American Psychiatric Association, 2022, p. 469). Ultimately, Social Cognitive Theory illustrates how ED behaviours and symptoms develop through the interaction of a child's thoughts, behaviours, and environment. Hence, interventions targeting a child's social cognitions, social skills, and the behaviour modelled in their environment can influence key elements involved in the development of EDs.

Transactional Model of Development. This model takes social cognitive theory a step further by extending beyond a child' s observations, behaviours, and thoughts to include how their behaviours influence the responses they receive from others and the interplay that evolves from meaningful relationships over time. Parents of children with EDs tend to experience significantly higher stress levels than other parents (Theule et al., 2013). A difficult temperament- characterised by low effortful control, high activity, inattention, and negative emotionality- is a predisposing trait strongly associated with EDs (Holmes et al., 2001). This, in turn, can elicit tougher responses, further reinforcing oppositional behaviour, and diminishing the effort parents put into teaching their child (Maccoby, 1984; Rothbart & Bates, 2006; Serbin et al., 2015). Children of any gender with EDs often show lower levels of positive parenting and increased mother-child conflict, which tend to

reciprocally reinforce each other over time (Georgiou & Fanti, 2014; Serbin et al., 2015). In classroom settings, children with EDs can be challenging to teach, their learning is affected, and their relationships with peers are negatively impacted (Döpfner et al., 2004; Lewis et al., 2017). EDs left unchecked are prone to a negative spiral where disruptive behaviours are met with harsher or stricter responses. In this transactional model, EDs do not develop in a vacuum, solely within the child. Instead, evidence demonstrates a reciprocal, bidirectional influence between the temperament and behaviours of children with EDs and the warmth or harshness of others' responses. When untreated, EDs are susceptible to a negative spiral where disruptive behaviours are often countered with harsher responses. Consequently, interventions should not aim solely to change the child. Instead, focusing on key individuals in a child' s life can also affect the developmental course of EDs. Supporting and enhancing the resilience of these caregivers to interact positively with children exhibiting challenging behaviours, along with teaching conflict resolution skills to nurture warmth in relationships, can help to interrupt negative transactional spirals.

The Ecological Systems Theory. This provides a broader perspective on the various relationships that influence a child's life, ranging from within the child, to microsystems of those closest to the child, to macrosystems of their wider social and cultural environment. Recognising the bidirectional influences involved in the developmental pathway of EDs, this theory encourages us to step back and examine specific subgroups at play in a child's development. The Dunedin longitudinal study has shaped the world's understanding of the interplay between genetic (nature) and environmental (nurture) factors (Poulton et al., 2015, p. 688). Multiple systems are known to have an impact on the development of ED, ranging from internal genetic and physiological factors such as a low resting heart rate, reduced functioning in arousal systems, and differing levels of neurotransmitters, to maladaptive relationships, deviant peers, and external factors such as poverty . Children who are the most vulnerable to developing problems in adolescence are often those who face multiple disadvantages across these systems (Gluckman & Hayne, 2011).

This theory has helped underscore the need to provide interventions across multiple systems to impact the various facets involved in the development of EDs. The influence of the field of developmental psychopathology and general psychological theories of development on this review is outlined in Figure 1. The diversity of risk and protective

factors within the child, from others and from relationships, emphasises the need to consider both individual and transactional outcomes. Furthermore, from an ecological theory perspective, the diversity of areas in which influence could occur highlights the importance of reviewing the mechanisms each intervention used to effect change, the types of outcomes targeted for change, and the subgroups each intervention aimed to influence. Conducting a review with a broad spectrum of intervention modes and mechanisms allowed for a comparison of how well diverse types of interventions aligned with developmental theories related to EDs.

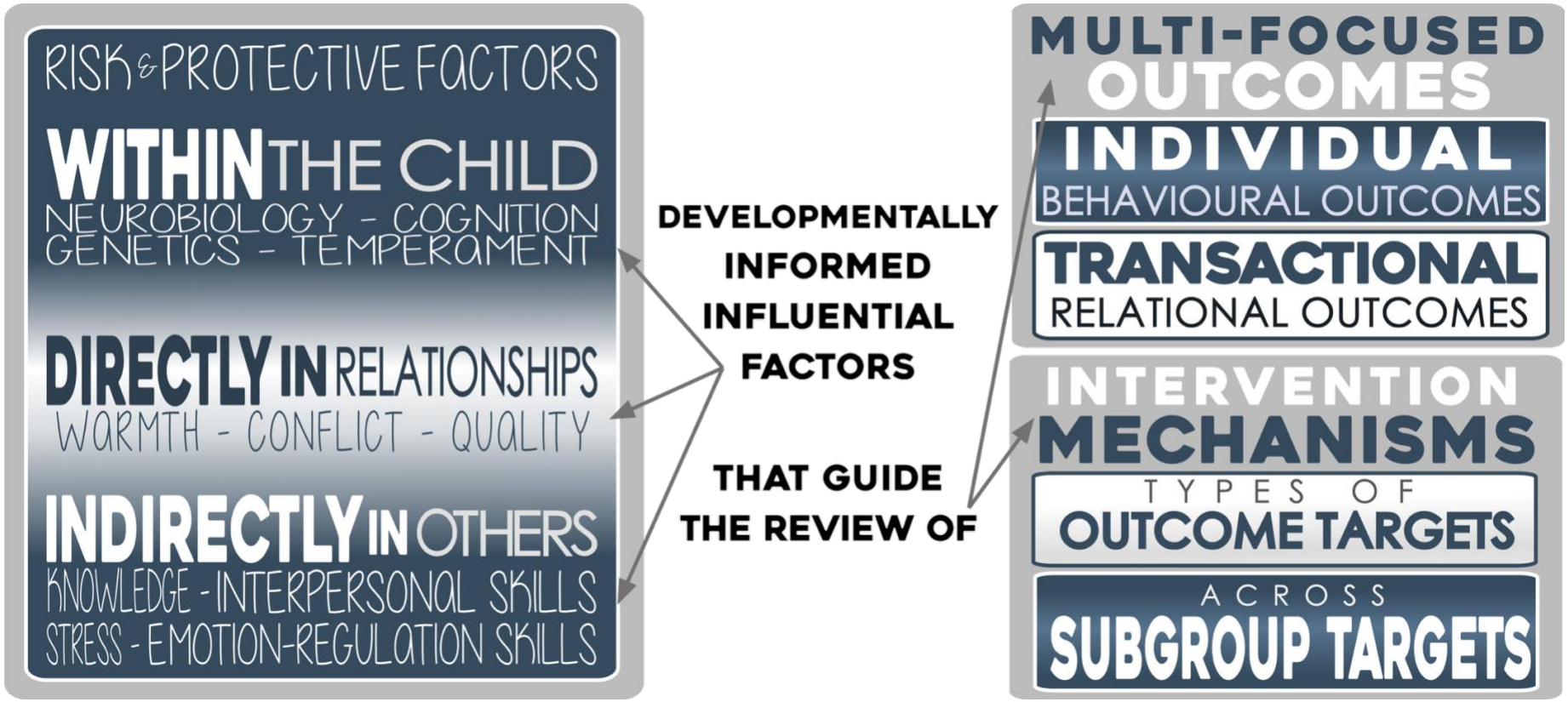
Review Importance

The current status of unmet treatment for EDs in Aotearoa, alongside identified areas of need and responsive recommendations outlined in ERO's report, all require evidence-based information relevant to EDs within the context of Aotearoa. Primary needs include addressing and enhancing the learning and professional development of schools and teaching staff on evidence-based interventions for challenging behaviour, addressing the lack of access to needed mental health experts, and filling a gap among current SLR reviews that can compare a broad range of interventions targeting subgroups across ecological systems in an evidence-based manner.

Schools and Staff Learning and Development. The Behaviour in Our Schools report clearly outlines recommendations to increase knowledge of evidence-based, effective interventions and strategies for managing challenging behaviours, building awareness and education on best practices in schools, and having an evidence-based foundation for developing accredited professional learning in behaviour management to boost

Figure 1

The Influence of Developmental Psychology on the Review Focus



teacher competency. Recommendation 7 is to "support schools to adopt evidence-based practices that promote positive behaviour and increase consistency of how behaviour is managed within the school" (Education Review Office, 2024, p. 122). Recommendation 15 is to "provide clear guidance to schools on what the most effective consequences for challenging behaviour are and how to use them to achieve the best outcome for students" (Education Review Office, 2024, p. 123). Recommendation 10 is to "prioritise evidence-based professional learning and development for teachers on effective approaches to manage behaviour and consider nationally accredited professional learning and development" (Education Review Office, 2024, p. 123).

A key element here is that government agencies are stakeholders who require evidence-based literature to inform policies and practices, as entire populations are affected by their decisions. Agencies funded *by* taxpayers must justify their choices, particularly in the fields of public health, education, and child development. Evidence-based information helps ensure that programs implemented are effective and minimise risk to the public.

My review addresses these needs by evaluating the effectiveness of evidence-based interventions related to the specific behaviours outlined in ERO's report. Including a wide range of interventions with different targets and modes of delivery allows for comparison of data across various interventions. I also provide additional information on interventions that can assist schools and communities in adopting evidence-based practices by not only reporting on the effectiveness of interventions but also outlining the mechanisms behind them.

Furthermore, SLRs possess a level of empirical quality that is suitable for government stakeholders. I adhere to strict pre-specified guidelines to minimise bias in my review and to ensure that the data extracted originates from high-quality empirical studies with rigorous research and reporting standards (Nightingale, 2009). Overall, my review offers relevant and usable information that will support ERO's recommendations.

Response to Lack of Access. A shortage of clinical staff and mental health experts, along with the resulting limited accessibility, is considered a crisis in Aotearoa, and with an aging mental health workforce, this situation is expected to worsen (Health New Zealand, Te Whatu Ora, 2024; Manatū Hauora, Ministry of Health, 2024; Mulder et al., 2023). In light of this, the reported rise of challenging behaviours in schools, the increasing proportion of

time spent responding to these behaviours, and the limited capacity—both in terms of time and expertise—to manage them are understandable (Education Review Office, 2024, p. 3).

My review provides an important and timely response to help reduce (though not eliminate) the challenge of accessing expert support. I do this by reviewing interventions that can be implemented and delivered by staff, regardless of their qualifications, and where training for intervention delivery is required, this training is also available to anyone, regardless of qualifications. This review reports on the effectiveness of interventions that are applicable and practicable in the context of Aotearoa.

Filling a Gap. Among SLRs and meta-analyses on interventions for externalising behaviour, a gap persists. Despite numerous factors as identified to influencing the development of EDs, an ecological approach is considered best practice for the treatment and prevention of EDs, employing one or multiple interventions to address various systems (Boden et al., 2010; Samek & Hicks, 2014). However, available reviews often focus on a single setting, modality, intervention type, or diagnosis and omit comparisons across various interventions targeting influential single factors and multiple factors. For example, reviews frequently concentrate on interventions with a single target subgroup, such as parent-based interventions (Mingebach et al., 2018), or a single modality, like teacher-delivered programmes (Aldabbagh et al., 2024), or one type of intervention, such as CBT alone (Riise et al., 2021), or focus on a single diagnosis, like ODD (Bradley & Mandell, 2005). The length of that sentence underscores how research on this topic is widespread with varying narrow foci. Consequently, current reviews fail to provide a comprehensive comparison of intervention effectiveness across ecological systems, modes, single, and multi-system approaches, or across a range of disorders related to EDs. This complicates the process of identifying best practices for intervention given the complexities involved in the development of EDs.

This review alternatively encompasses broader eligibility criteria to provide a complete overview of various interventions available for EDs. The interventions included are intentionally wide, reflecting current theories on ED development and the multiple systems involved. From here, readers can explore more specific paths to identify best practices within categories of interventions for EDs; however, this initial comprehensive overview is missing in the literature. Additionally, including a broad range of evidence-based

interventions for EDs may help different communities identify best practices suited to their resources and context.

Findings From Similar Reviews

Interventions considered most effective in comparable reviews generally target younger children, involve parents, or encompass multiple systems. Mingebach et al. (2018) and Buchanan-Pascall et al. (2018) report moderate and consistent effects for parent-based interventions. Although long-term data are limited, teacher-delivered interventions demonstrated small to moderate effects, particularly with younger children, and provided some evidence of enhancing student-teacher relationships (Aldabbagh et al., 2024; Franklin et al., 2017). Programmes integrating interventions with parents, teachers, and children exhibited significant effects, and combined models of care were recommended (Baumel et al., 2021; Sawyer et al., 2015). In severe cases of antisocial behaviour, Tan et al. (2017) identified multisystemic therapy as particularly effective. Studies on personalised interventions indicated some short-term improvements; however, empirical support remains limited (Skinner et al., 2024). Given that each review concentrates on a specific type or target of intervention, perceptions of effectiveness may be influenced by feasibility bias, methodological convenience bias, or evidence accessibility bias inherent to each intervention's characteristics. Conversely, my review endeavours to compile studies of interventions with varied modes and modalities into a comprehensive overview, thereby facilitating consistent and thorough comparisons across them.

Effectiveness Review

This is an intention-to-treat (ITT) review involving broad interventions that are suitable for implementation in both public and private settings. These interventions target any or all subgroup(s) across an ecological systems framework. Due to the psychological focus of this thesis, interventions reviewed included cognitive, behavioural, psychosocial, skills-based, and psychoeducational approaches aimed at children exhibiting EDs or related subgroups.

Including a wide range of intervention types meant the purpose of this review was descriptive rather than a meta-analysis. However, narrative reviews are vulnerable to bias due to less standardised procedural methods. To ensure objectivity, this review used a

systematic methodology combining textual narrative synthesis and metasummary approaches. Standardised methods were applied for data extraction, the categorisation of included studies and outcome measures, bias and certainty assessments, and the synthesis and reporting of results (Xiao & Watson, 2019).

Outcome measures were categorised based on ED symptomology related to behaviours mentioned in the ERO report and influential factors noted in developmental theories of EDs. Effectiveness was evaluated across both homogenous (e.g., similar constructs measured across studies) and heterogeneous subgroups (e.g., differences in intervention mechanisms and subgroup targets). This approach enabled the reporting of the effectiveness of specific intervention types and the reporting of particular moderators, guided by developmental theories, which highlighted influential factors over effectiveness as well as areas of strength and weakness in the existing research. I outlined the process of systematic procedures undertaken in a logic model for clarity (see Figure 2).

Objectives

The primary objective of this thesis was to conduct a comprehensive and systematic review of the literature concerning the effectiveness of evidence-based interventions for primary school children aged 5–14 presenting with EDs. Specifically, this review seeks to:

1. Offer a thorough overview of interventions for EDs across diverse targets, settings, and modalities to address gaps in existing reviews, whilst ensuring that included interventions do not depend on psychologist involvement, given the limited availability of psychologists in Aotearoa.

2. Provide insights into current challenges highlighted in the ERO report I referenced, with particular regard to schools in Aotearoa, thereby supporting evidence-based practice and professional development.

3. Clearly communicate the certainty of evidence to prevent mere regurgitation of study results and to reduce potential misinterpretations of the evidence reviewed.

4. Investigate the moderating factors of intervention and study design that influence the effectiveness of measured outcomes. This aims to identify areas of strength, weakness, saturation, and gaps within the reviewed results, especially considering developmental theories of EDs and rigorous study methodologies.

5. Ensure that my review remains pragmatic and accessible to a broad audience. Results should be presented in formats that facilitate the identification of specific intervention types or target behaviours, while minimising the use of excessive jargon.

Method

My review adhered to PRISMA guidelines (Liberati et al., 2009; Page et al., 2021) and was registered with the Open Science Framework (OSF), along with records of search strategies, screening processes, and result syntheses https://osf.io/9afzx/ (Powell, 2024). As this was not a meta-analysis, to improve replicability, I used a hybrid of textual narrative synthesis and metasummary approaches (Xiao & Watson, 2019), ensuring that data extraction, bias and certainty assessments, and synthesis procedures were conducted systematically and guided by validated or recommended instruments. I adhered to SWiM guidelines for synthesis and GRADE for certainty ratings (Campbell et al., 2020; GRADE Working Group, 2004).

Eligibility Criteria

Eligible studies needed to be published in peer-reviewed journals available in full text and written in English. These restrictions aimed to ensure the review included evidence-based literature. Studies failing to meet these criteria were excluded. Moreover, eligibility criteria were guided by the PICO framework (Population, Intervention, Comparison, Outcome), which is especially suitable for comparing interventions (Huang et al., 2006). I developed a prescriptive PICO eligibility guide to facilitate objective and replicable decision-making during study selection, screening, and data extraction (see Appendix A).

Population

The target population was primary-aged children displaying EDs, along with subgroups who regularly interact with these children, such as teachers, peers, or guardians (termed as parents throughout the rest of this review).

Age. Primary age was defined as 5–14 years, aligning with the legal age range for primary school attendance in Aotearoa (Parliamentary Counsel Office, 2020). Studies were included if the mean age of participants, including standard deviations from the mean, fell within this range. For example, if the mean age of participants was 13 but the standard deviation encompassed 2 years either side of the mean, the study would be excluded. This

ensured that included studies would be weighted toward primary-aged children rather than toward the target age range limits, which are more relevant to teenagers and toddlers. Although most students finish primary school by age 11, the broader range was chosen to provide a review that is inclusive of students who remain in primary school longer and to align with legal definitions. Studies with population means outside this range were excluded.

Externalising Disorders. Eligible EDs were only those diagnosable within the target age range and classified using the HiTOP model (Kotov et al., 2017). Classifying EDs by HiTOP meant searches extended beyond diagnosable disorders to include symptomology relevant to EDs, while still adhering to an evidence-based taxonomy (Ringwald et al., 2023). Eligible ED disorders under HiTOP included ADHD, IED, CD, and ODD, as well as symptoms characterised as disruptive or antisocial. Utilising HiTOP allowed an eligibility criterion aligned with the specific behaviours outlined in the ERO report (ERO, 2024, pp. 16-17), enabling a broader inclusion of relevant studies.

Subgroups. Subgroups associated with the target children were incorporated to encompass interventions aimed at key individuals within the child's social environment, such as parents, teachers, and peers. This methodology aligns with developmental frameworks, including Bronfenbrenner's ecological systems theory and the transactional model of development, which acknowledge the bidirectional influence between children and their surroundings (Bronfenbrenner, 2000; Sameroff, 2009). The inclusion of subgroups also mitigates an overly individualistic perspective of EDs, which may unjustly attribute responsibility and causality solely to the child (Gatti et al., 2018). Moreover, the integration of subgroups related to target children supports current understandings of the influential interaction that occurs between nature and nurture in child development (Stiles, 2011).

Intervention

Eligible interventions could target children, the physical environment, as well as multiple subgroups related to target children, such as parents, teachers, and peers. Interventions were not limited by delivery mode. Removing restrictions on intervention type and delivery mode aimed to first, fill a gap across current SLRs, which have narrowly-focused inclusion criteria and do not comprehensively review the field – making comparisons across interventions difficult, and secondly, to enable evidence-informed decisions relevant to

people from varying contexts. Interventions had to be feasible within school/public and home/private environments and needed to account for the low supply of clinical psychologists and mental health professionals in Aotearoa (Burns & Reporter, 2024; Health New Zealand | Te Whatu Ora, n.d.). Eligible interventions were behavioural, psychosocial, psychoeducational, or multisystem-based approaches, which could be delivered by easily accessible staff. Specifically, only those requiring resources and professional staffing that are reasonable and readily available were included. Interventions involving medication, dietary changes, sleep modifications, or requiring delivery by high-level mental health professionals were excluded.

Comparison

Randomised controlled trials (RCTs) featuring a sufficient randomisation process with balanced and clearly isolated control groups were included to review the intervention effects as compared to the typical developmental trajectory of like peers and minimise confounding variables (Sameroff & Mackenzie, 2003). Eligible control conditions included treatment as usual, business as usual, services as usual, waitlists, no treatment, or attentional controls – notably, all 'as usual' controls were all labelled as business as usual (BAU) in this review for clarity. Multi-arm designs were acceptable if each experimental group was individually compared with the control group. Direct intention-to-treat (ITT) statistical analyses between one experimental group and one control group were required to evaluate the efficacy of interventions. Studies comparing interventions to other interventions were not included. The effectiveness of an intervention compared to a control must be established first. Critically, included studies had to report the interaction effects between groups (experimental and control) across time (from pretest to posttest) – that is, Group × Time effects (G×T). G×T analyses were necessary in determining whether differences in the effects found between groups over time were attributable to the intervention rather than to chance or external factors. Reviewing effectiveness without G×T calculations risks misrepresenting whether effects found were attributable to interventions in ITT reviews (Twisk et al., 2018). Therefore, studies omitting G×T interaction effects were excluded.

Outcome

Eligible outcomes were guided by behaviours described in the ERO report, the HiTOP model and risk and protective factors associated with the development of EDs (Bandura,

1986; Bronfenbrenner, 2000; Education Review Office, 2024; Kotov et al., 2017; Sameroff, 2009). Eligible outcomes represented individual, interpersonal, transactional and environmental factors. Eligible individual outcomes included measures of target children's anti- and prosocial behaviours and symptoms related to externalising or disruptive behaviours. All individual outcomes of target children were categorised in the behavioural domain in the result syntheses. Eligible interpersonal outcomes included measures of closeness, warmth, or conflict between target children and related subgroups. Eligible transactional outcomes included measures of related subgroup factors that impact relationship quality, such as parenting styles, knowledge, or stress levels. Interpersonal and transactional outcomes were categorised in the relational domain in the result syntheses. Outcomes measuring academic achievement were excluded, as a child's academic attainment is not necessarily indicative of a child's disruptive behaviour or prosocial engagement. Children may be highly engaged yet achieve poor grades, and vice versa. Furthermore, cognitive psychometric measures of attention or short-term memory in a clinical setting were excluded, as there was insufficient evidence to support their external validity in identifying ADHD symptoms (Hall et al., 2016; Nichols & Waschbusch, 2004). An exception for the inclusion of cognitive outcomes was measures using the Behaviour Rating Inventory of Executive Function (BRIEF), as this instrument was designed explicitly for natural contexts and showed good ecological validity (Qian et al., 2017a).

Information Sources and Search Strategy

I undertook searches, commencing on 18th September 2024. Databases searched included Index New Zealand (INNZ), Proquest, ScienceDirect, Scopus, Wiley, and Waikato University's Library search engine, Discovery. Discovery included the following databases: APA PsycArticles® (Ovid); APA PsycInfo® (Ovid); APA PsycTherapy® (Ovid); Cochrane Library (Wiley); Counselling and Therapy in Video Library (ProQuest); INNZ; ProQuest Social Sciences (ProQuest); PTSDpubs (ProQuest); ScienceDirect (Elsevier); and Scopus (Elsevier).

The search terms used to identify eligible studies were Child*, Behaviour, Effect, Treatment OR Intervention, Random, Externalising OR Disruptive OR ADHD OR "Oppositional Defiant Disorder" OR "Conduct Disorder" OR "Intermittent Explosive Disorder" OR Antisocial. Asterixis and speech marks were used with search terms to include variations of word endings and to clarify specific diagnoses, respectively. With various broad terms often used

for the same concept, I concentrated the majority of search terms on describing EDs (Bayliss & Beyer, 2015). I also reran the searches to alternate between the terms Treatment and Intervention. Search terms used to exclude ineligible studies using the "NOT" Boolean operator were Meta-analy*, Systematic, Preschool, Kindergarten, Infant, Medic*, Stimulant, Dose, Pharmacotherapy, Internalising, Depression, Anxiety, and Telehealth.

General "Full Text", "Peer Reviewed", and "English" filters were applied across databases as available. No date limitation was applied. These filters enabled the identification of eligible evidence-based articles, regardless of their publication date. My entire search process, including dates, URLs, variations of search terms and the application of filters across databases, was documented in the Line-by-Line Search Process document (see Appendix B).

Selection Process

Search results from all databases were exported into an Excel file titled "Collated Search Results", with separate sheets added for each stage of the screening process and were uploaded to the OSF. First, duplicates were screened and removed. Two reviewers then independently evaluated all titles and abstracts using my prescriptive PICO eligibility guide document, which was also uploaded to the OSF. Interrater agreement was 82.8% with interrater reliability showing substantial agreement bordering on almost perfect (Cohen's kappa = 0.72) before discussions of discrepancies (Landis & Koch, 1977). Interrater reliability was determined on the recommended 25% of the sample (Armstrong et al., 2020; Belur et al., 2021). All discrepancies were discussed and resolved without requiring a third reviewer. I then proceeded to examine the full texts of all remaining articles against the prescribed PICO eligibility criteria.

Data Collection

The Cochrane group recommends tailoring data collection to suit individual review (Cochrane Methods, 2016). To determine which data items were required for my review, I consulted several validated tools to identify how I would report results, assess bias, synthesise results, and conduct sensitivity analysis and certainty of evidence assessment.

I used the Cochrane "Data Collection Form: Intervention Review – RCTs only" to identify recommended study, intervention, and outcome data items to collect (Cochrane Methods, 2016). I also referenced the Template for Intervention Description and Replication

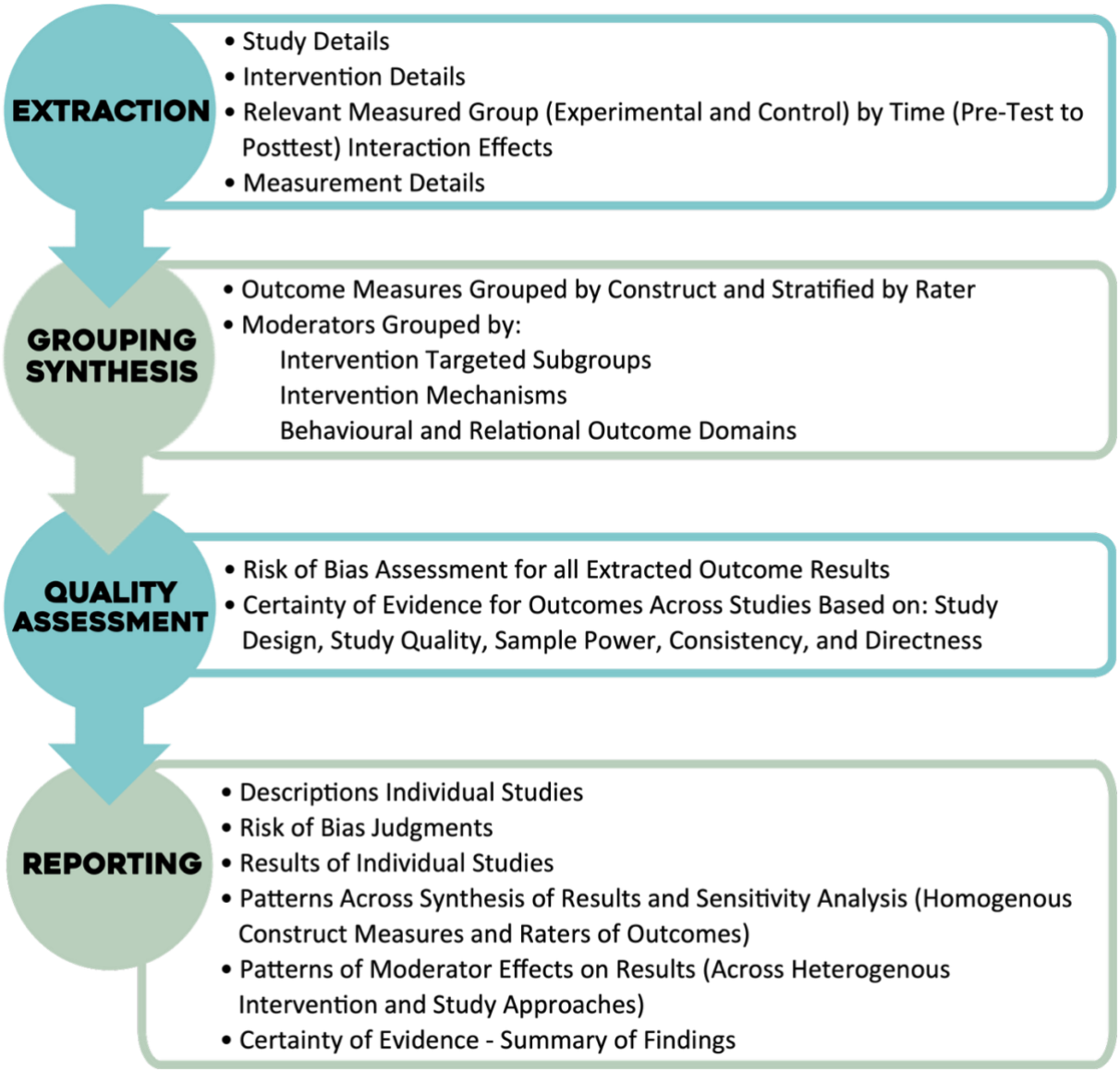
(TIDieR) to identify recommended intervention characteristics for extraction. TIDieR was explicitly designed to improve the replicability of data collection in intervention-based reviews and ensure the data items collected provide useful information, which supported my thesis objective of delivering clear and applicable information in my review (Hoffmann et al., 2014). I used the Synthesis Without Meta-analysis (SWiM) reporting guideline as it is recommended for use alongside PRISMA to improve the replicability of synthesis procedures (Campbell et al., 2020). SWiM guided the grouping of outcomes and other extracted data for synthesis, thereby informing the data items that required extraction. Finally, I referenced the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) clinical guidelines to inform the selection of data items for assessing the certainty of evidence (GRADE Working Group, 2004; Guyatt et al., 2008). Figure 2 depicts a logic model of the systematic processes involved in handling data.

Data Items

Study details, intervention characteristics, and reported outcomes were collected and compiled into an Excel dataset in long format, where each new row represented one included outcome, along with its related data items. This long-format structure facilitated later synthesis by enabling individual outcomes to be easily filtered or grouped across other extracted data fields. The dataset was stored in a master spreadsheet titled "Master Data Long Formatting Powell" and was uploaded to OSF.

Figure 2

Logic Model of Systematic Procedural Steps



Study details extracted included: ID, Reference, Comparator, Country, Target child age (minimum maximum, and range), Characteristics of target children, RCT design (individual or cluster), Study sample size, Sample size breakdown (number of children, teachers, parents, classrooms, schools involved), Sample power (low or sufficient), and Number of extracted outcomes per study.

The binary reporting of sample power as either low or sufficient was based on the study authors' judgments and used for clarity, as authors varied in their descriptions of sample power. This binary categorisation was used to provide a standardised and simple

expression of the certainty of evidence calculations, contextualising reported outcomes (Boon & Thomson, 2021).

Intervention details extracted included: Intervention name, Intervention target population (single, dual, or multisystem), Intervention subgroup targets (target children, peers, parents, teacher), Certified intervention trainer required (yes, no), Certified intervention facilitator required (yes, no), Intervention Facilitator (Who Implements Intervention), Intervention mechanisms used, and Number of mechanisms used per study.

The certified trainer and facilitator data items were included to identify interventions that required programme-related certification. Certified trainers refer to interventions that require certified individuals to train others in programme facilitation. Certified facilitators refer to interventions that require certified individuals to facilitate the programme.

Outcome details extracted included: GxT result page number, GxT result table number, Outcome name, Measurement method (Scales, Observations, Interviews, Coded observed materials, etc), Outcome Rater (Who or what provided ratings/measures), Data Type (continuous or binary), Reported GxT interaction effect rounded to 2dp, Reported standard error, Reported degrees of freedom, 95% Confidence intervals, p-value < .05, Direction of effect (Beneficial, Adverse), Effect Size Category (Trivial, Small, Moderate or Large), Effect Size Rounded 2dp, Outcome Domain (Behavioural or Relational), Outcome Domain Subcategory (Indirect, Direct, Reduce, or Improve), SWiM Primary Construct Domain, SWiM Primary Construct Subgroup, and SWiM Secondary Rater.

Studies with missing data were included in my review; however, this limitation restricted my ability to consistently report effect sizes and confidence intervals across the included studies. Therefore, confidence intervals were not included in the data synthesis, and only available effect size descriptions were reported.

Separate sheets were added for columns that contained multiple data entries per cell and were entered separately in long format. This way, all data was readily available for synthesis and reporting. My synthesis methods detail the data items entered for mechanisms, outcome domains, and SWiM constructs.

Study Risk of Bias Assessment

The risk of bias (ROB), otherwise known as systematic error, was assessed for each included outcome using the revised Cochrane Risk of Bias tools for individual (ROB2) and

cluster (ROB2C) randomised trials to conduct the risk of bias assessments (Sterne et al., 2019). Each outcome measure was assessed using the signalling questions for cluster and individually randomised studies found in the publicly available Excel templates developed by the Cochrane group (Sterne et al., 2019a, 2019b). I made no alterations to these standardised tools. Bias judgments were categorised as low concern, some concern and high concern – representing the believed probability of introduced bias. I completed ROB judgments with supporting evidence for all domains relevant to ITT analyses, including the overall judgment (OJ). As a sole ROB assessor, I noted every rationalisation with page references to facilitate cross-checking of my judgments. A record was kept for all completed ROB assessments and was uploaded to OSF.

Low to high judgments for all outcomes were collectively presented in a forest plot across each domain and the OJs. Low-to-high judgments for each outcome across all domains were also presented in a traffic plot. ROB judgments and justifications, as well as the impact of any introduced bias, were also included in the master data spreadsheet to be used in result synthesis, sensitivity analysis and judgments made on the certainty of evidence.

Effect Measures

Being a hybrid of a textual narrative synthesis and metasummary, effects measures were categorically rather than numerically standardised. Direction-of-effect, as defined by reported GxT interaction effects for each included outcome, were used as the standardised effect measure in my review. A binary direction-of-effect was reported as 'Beneficial' or 'Adverse', which was determined based on whether GxT effects showed a beneficial or adverse impact on the measured outcome. Standardising effect measures by the direction of effect enabled the reporting of intervention effectiveness across multiple types of outcomes. All GxT effects were included irrespective of statistical significance because using p-value cut-offs to determine included effect measures is criticised as showing poor transparency and being statistically misleading due to the varying power to detect effects across included studies (Higgins et al., 2022; Valentine et al., 2010). However, including all effects irrespective of significance can be misleading, as it does not show the level of confidence that can first be placed in the studies' capacity to detect an effect, nor the level of confidence that can be placed in whether effects are attributable to the intervention rather

than a confounding factor. To support my objective of providing evidence-based information which is not only regurgitated but considered alongside the quality of evidence, especially if my review is used to inform the use of interventions with children, direction-of-effect was reported alongside contextual information such as statistical significance, sample power and effect size (Boon & Thomson, 2021; Makin & Orban De Xivry, 2019; Schünemann et al., 2020; Thomson & Thomas, 2013).

GxT effects that were statistically significant were indicated alongside direction-of-effect using a sign test as recommended by Thomson and Thomas (2013). GxT effects were considered statistically significant at $p < .05$. as a commonly accepted p-value threshold for reducing Type I errors (false positives) (Makin & Orban De Xivry, 2019). This indicated what effect measures had statistical confidence that the effects found were attributable to the intervention. The sample power to detect statistically significant effects was also reported alongside direction-of-effect, as low or sufficient (as reported by study authors). If an effect measure was associated with sufficient power this indicated that it came from a study with sufficient capacity to detect statistically significant effects. Non-significant effect measures associated with low power was not necessarily indicative of an insufficient intervention, but that the effect measure was associated with a study that was underpowered to detect statistically significant effects in the first place. Although many studies did not report effect sizes, this is considered valuable for understanding intervention effectiveness (Higgins et al., 2022). Therefore, all available effect sizes were included and reported as trivial (not reaching the recommended cut-off for small), small, moderate, and large (Cohen, 1992).

Reporting effect measures by direction-of-effect alongside statistical significance, sample power, and effect sizes available enabled me to address my overall objective to review the range of evidence of effectiveness across interventions. Furthermore, it enabled me to meet another objective of providing a review that considers the quality of evidence behind reported effects so as not to blindly regurgitate reported effects. The clear categorical nature of reporting effect measures also seemed the best approach to meet my objective of providing a review with accessible and clear information enabling a wide range of readers to make informed choices on implementing interventions.

Synthesis Methods

Due to the comprehensive coverage of interventions and the range of outcome measures included, a meta-analysis was not used (Ioannidis, 2008). However, narrative-based reviews are criticised for being vulnerable to bias due to using less transparent and systematic methods, and allowing selective reporting by review authors (Xiao & Watson, 2019). To enhance transparency and replicability, I followed SWiM items to guide the synthesis of quantitative data without meta-analysis, where systematic and standardised methods were included on top of a traditional narrative approach, which allowed me to provide an evidence-based review without meta-analysis.

To review the effectiveness of interventions for children exhibiting externalising behaviours, I synthesised results by grouping measured effects into similar construct domains and then reporting the direction-of-effect using vote counting – an acceptable method for non-meta syntheses (Higgins et al., 2022). Synthesised results were reported alongside certainty ratings based on the quality of evidence using the GRADE framework (Bezerra et al., 2022; GRADE Working Group, 2004; Guyatt et al., 2008).

Outcome Groupings

In following SWiM guidelines, the heterogeneity of included interventions and the types of outcome measures was accounted for by identifying homogeneous constructs across measures, which allowed for varying effects across varying interventions to be described in a consistent manner (Campbell et al., 2020; Higgins et al., 2022).

Outcomes were grouped using a dual-layer framework. Outcomes were primarily grouped by construct domain guided by the functional domains outlined in Dodd et al.'s (2018) 38-item taxonomy, which was further categorised into subgroups within each construct domain. Construct was selected as the primary grouping of outcomes as it reduced the heterogeneity of reported outcomes and improved comparability of effects found across differing interventions. Construct domains provided a simple overview of intervention effects, and construct subgroups provided greater details of the types of effects.

The second layer stratified construct groupings by measurement raters to support easy identification of any effect patterns between raters. Rater stratification was included due to the high level of unblinding across included studies. This dual-layer grouping aligned with SWiM guidance to group studies for synthesis based on outcome and study design to

improve interpretability of reported results (Campbell et al., 2020). SWiM groupings were recorded in the Master Data Long Formatting Excel spreadsheet and uploaded to OSF.

Valentine et al. (2010) posed a relevant critique of review authors who do not use meta-analysis due to heterogeneity, only to later report an overall sweeping summary across the evidence. Therefore, reported results referred to these homogeneously grouped outcomes.

Moderators

Moderators in the results synthesis were categorised by intervention mechanisms, intervention target subgroups, and relational and behavioural domains of the intervention. The proportion of moderator use and the patterns of impacts moderators had on effects were reported and outlined in tables and figures. Grouping outcomes by these moderators provided readers with quick access to information on specific types of interventions relevant to them and enabled me to comment on any patterns across interventions and studies.

Mechanisms

Mechanisms reflect the theoretical agent of change utilised by each intervention. Some interventions use multiple mechanisms, while others only use a couple. Mechanisms included were: Behaviour Tracking, Psychoeducation, Self-Management Skills Training, Attention Training, Learning Support, Behaviour Management Training, Social Emotional Learning, Multisystem Collaboration, and Physical Environment Alterations. This moderator highlighted how interventions theoretically intended to effect change, making it possible to compare what mechanisms were most used and the extent to which they aligned with risk and protective factors outlined in developmental theories.

Target Subgroups

As mechanisms highlight *how* interventions intended to effect change, identifying the target subgroups of each intervention highlights who each intervention considered requiring an intervention to affect change. This moderator enabled me to compare the heterogeneity of targets across interventions and how these align with ecological systems theory and the transactional model of development.

Relational and Behavioural Domains

These domains highlighted the types of outcomes interventions targeted and were valid measures of effectiveness for each intervention. I reviewed the patterns of outcome

types for areas of saturation and gaps, which allowed me to make comparisons with theoretical causes and symptoms of externalising disorders in developmental theories.

Sensitivity Analysis

I included a comparative table and figure of results where outcome measures with a high risk of bias and low sample power were removed, leaving outcome measures with high statistical confidence in the reported effects. I included this to demonstrate the volume of high-quality, evidence-based intervention effectiveness available.

Certainty of Evidence

A modified GRADE assessment of outcomes, guided by the GRADE Working Group (2004), was used to evaluate the quality of evidence included in this review. Factors supporting assessment included: Study Design, Study Quality, Sample Power, and Directness. For Study Design, all included studies were RCTs, so all were rated as high. Study Quality assessed study methods and execution, and ratings utilised the ROB judgments (low, some, high concern). Sample Power assessed the studies capacity to detect effects and were rated as low and sufficient. Consistency was assessed by evaluating how similar the outcome measures were across the included studies. Ratings considered whether outcomes within SWiM construct domains were majority consistent with minimal outliers, or if outcomes were highly diverse. Directness assessed the similarity between included outcomes and outcomes of interest. Ratings of directness considered the similarity of included outcomes to the target age range, and the behaviours outlined in EROs review and the related disorders. Sample power to detect effects was categorically rated as low and sufficient.

Data Presentation

Individual study results were presented in a table and grouped by study, with results being grouped by the behavioural and relational domain and subcategory first, then by construct domain and construct subcategory, and finally by rater. Ordering results this way allowed consistent reporting from introduction themes to results synthesis. Furthermore, behavioural domains divided outcomes into those intended to be improved or reduced, making it clear to readers what was measured. Intervention characteristics and results were presented in separate tables.

The synthesis of results was presented in a tabular format, complemented by harvest plots illustrating vote-counting based on effect direction. These visualisations effectively demonstrated the intervention's efficacy, explicitly indicating whether effects were observed and if they were beneficial or adverse (Higgins et al., 2022). The harvest plots also symbolically conveyed information regarding sample power and statistical significance, thereby providing contextual insight into the confidence level of the reported effects. The table comprehensively detailed the measured outcomes, SWiM groupings, GxT effect reports, significance levels ($p<.05$), effect sizes, measurement methodologies, measurement raters, blinding status of raters, and indicators of evidence certainty. Outcomes were initially categorised by behavioural relational domains and subcategories, subsequently by construct domains and subcategories, and ultimately by outcome names in the presentation. This structured approach offered a clear overview of the outcome measures associated with each construct domain and type of measurement. Additionally, a summary of findings table was included, outlining the certainty of evidence and synthesised results.

Results

A total of 1,969 studies were identified through database searches. Of these, 497 duplicate studies were removed. An additional 1449 studies were excluded because they did not meet the detailed eligibility criteria outlined in the PICO framework guide, provided in Appendix A. The included studies focused on children and related subgroups, had minimal restrictions on who could deliver interventions, and reported sufficient results for ITT analysis. Twenty-three studies met the eligibility criteria and were included in this review. The screening process is detailed in the PRISMA flowchart (see Figure 3). Screening justifications were documented throughout the Collated Search Results Excel file, uploaded to my registered study on OSF (Powell, 2024).

Study Characteristics

All included studies were RCTs with a clear control group, conducted between 1995 and 2024. Population sample sizes ranged from 24 to 3084, with target populations including children, parents, classrooms, teachers, schools, and multisystems. Details of the included study characteristics are provided in Table 1.

Intervention Details

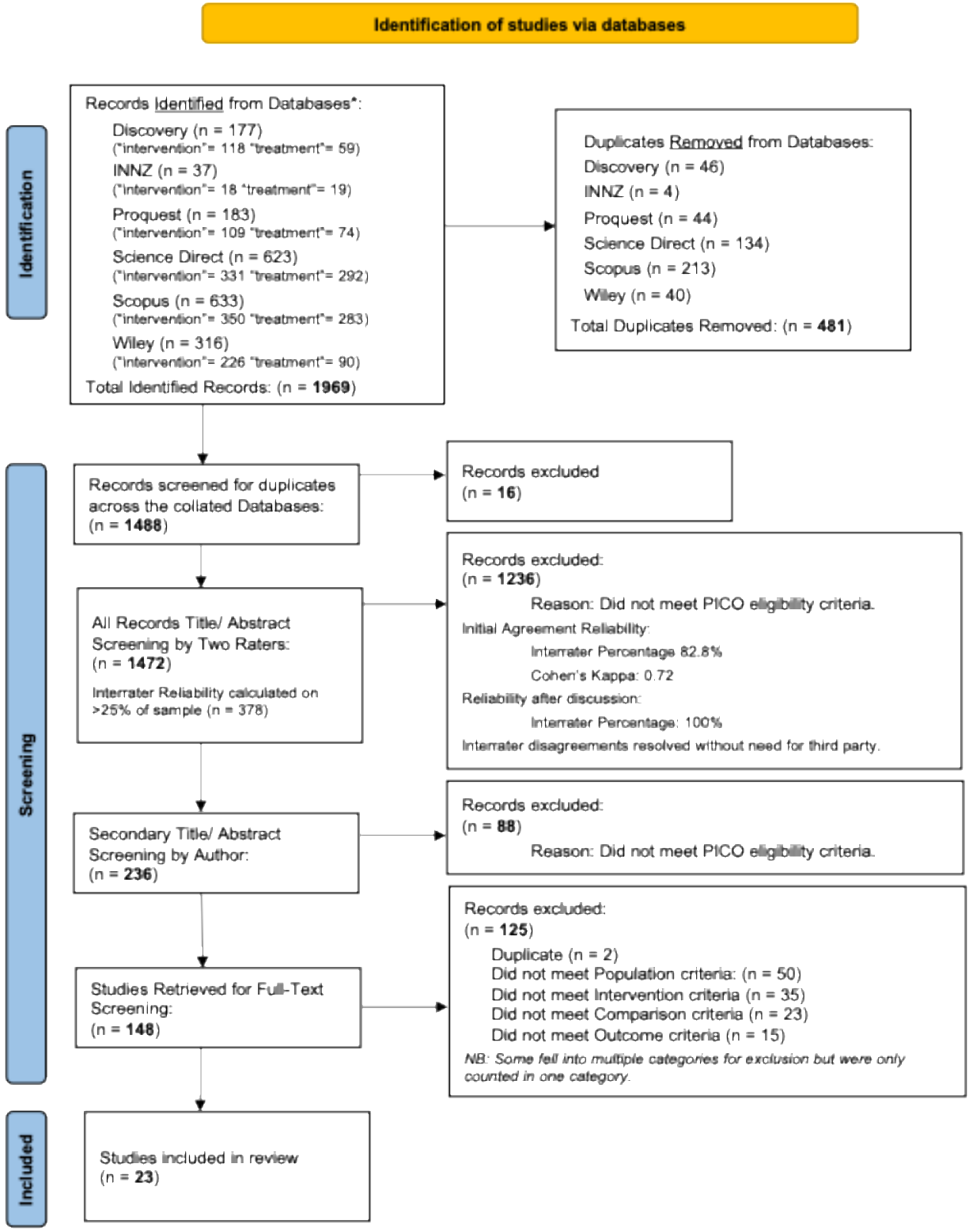
A summary table of intervention details outlines intervention targets, mechanisms of change used, and the type of training and facilitation needed (see Table 2). An outline of each intervention is given in the text.

The Positivity and Rules programme (PR)

PR (Veenman et al., 2017) offers teachers psychoeducation on ADHD and training in classroom behaviour management strategies, including setting clear rules and using positive reinforcement (both antecedent and consequent techniques) over 18 weeks. PR, together with the added use of a daily report card, is available at three different levels of intensity for individual children with ADHD symptoms or emerging challenging behaviour.

Figure 3

PRISMA Flowchart



Note. Adapted from Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372,* Article n71. Licensed under Creative Commons CC BY 4.0.

Table 1

Study Characteristics

ID	Reference	Intervention Name	Abbreviated Name	Comparator	Study Size	Sample Power	Target Child Age	Target Child Characteristics	GxT Effects Located
23	Veenman et al. (2017)	The Positivity and Rules Programme	PR	BAU	114 Children; 87 Teachers	Sufficient	6 to 13	ADHD Symptoms	Page 7
27	Thompson (2014a)	Self-Management Training and Regulation Strategy	STARS	BAU	1089 Children; 42 Teachers; 7 Schools	Sufficient	11	Disruptive Behaviours	Pages 421 and 422
43a	(Kirk et al., 2022)	Tali Train (Placebo)	TALIP	Placebo	98 Children; 8 Classrooms; 3 Schools	Sufficient	5 to 9	Children	Tables 2 and 4
43b	(Kirk et al., 2022)	Tali Train (BAU)	TALIB	BAU	98 Children and 8 Classrooms across 3 Schools	Sufficient	5 to 9	Children	Tables 2 and 4
44	Seeley et al., (2009)	First Step To Success	FIRST	BAU	42 Children (sub-sample)	Sufficient	6 to 10	>5 ADHD Symptoms	Page 44
52	Cook et al., (2018)	Establish-Maintain-Restore	EMR	Attention	220 Children; 10 Teachers; 3 Schools	Low	10 to 11	Disruptive Behaviours; Teachers requiring improved relationships	Table 3
89	Chesterfield et al. (2021)	1-2-3 Magic	MAGIC	BAU	57 Parents	Low	6 to 12	ADHD	Table 3
92	Meany-Walen et al. (2014)	Adlerian Play Therapy	PLAY	Attention	58 Children	Sufficient	5 to 9	Disruptive Behaviours	Page 51
94	Klopfer et al., (2019)	Errorless Classroom Management	ECM	BAU	118 Preservice Teachers	Sufficient	5 to 16	Children	Tables 2 and 3

Table 1 continued)

ID	Reference	Intervention Name	Abbreviated Name	Comparator	Study Size	Sample Power	Target Child Age	Target Child Characteristics	GxT Effects Located
95	Ashworth et al.,(2020)	Good Behaviour Game	GBG	BAU	3084 Children across 77 Schools	Low	7	Children	Table 4
107	O'Hare et al. (2015a)	Mate-Tricks	MATE	BAU	347 Children; 8 Primary Schools	Low	9 to 10	Socioeconomic disadvantage	Table 5
128	Murray et al. (2008)	Daily Report Card and Parent-Teacher Consultation	DRC+	BAU	24 Children	Sufficient	5 to 11	ADHD and Classroom Impairment	Table 3
166	Hosseinnia et al.,(2024)	Psychoeducation	PSYCHED	BAU	72 Children; their Parents and Teachers; 12 Schools	Low	6 to 12	Children with ADHD	Tables 2 and 3
171	de Jong et al., (2023)	Self-Help Parenting	SELF	BAU	110 Families	Low	4 to 12	Externalising Behaviour, >2 ADHD symptoms and/or >1 ODD symptoms causing impairment	Figures 2 and 3
205a	Bradshaw et al., (2020)	GBG Only	GBGO	BAU	1526 Children; 318 Classrooms; 27 Schools	Low	5 to 11	Aggressive-Disruptive Behaviour	Table 3
205b	Bradshaw et al., (2020)	GBG and PATHS Promoting Alternative Thinking Strategies	PATHS+	BAU	1526 Children; 318 Classrooms; 27 Schools	Low	6 to 11	Aggressive-Disruptive Behaviour	Table 3
208	Hoogendijk et al., (2020)	Key2Teach	KEY	BAU	103 Teacher-Student Dyads; 44 Schools	Low	7 to 14	Externalising Behaviours	Table 4

Table 1 (continued)

ID	Reference	Intervention Name	Abbreviated Name	Comparator	Study Size	Sample Power	Target Child Age	Target Child Characteristics	GxT Effects Located
232	Qian et al. (2017b)	Ecological Executive Skill Training	EEST	BAU	68 Children	Sufficient	6 to 12	ADHD	Tables 2, 3 and 4
257	Sibley et al. (2013)	Supporting Teens' Academic Needs Daily	STAND	BAU	36 Children	Low	11 to 15	ADHD	Tables 3 & 4
828	Braun et al., (2020)	Seating Chart	SEAT	Attention	1573 Children; 59 Classrooms	Sufficient	10 to 11	Children	Tables 3 & 4
861	Cook, et al., (2018)	Positive Greetings at the Door	GREET	Attention	203 Children; 10 Teachers; 2 Schools	Low	12 to 14	Disruptive/Off-Task Behaviours	Table 1
988	Cook et al., (2017)	5:1 Positive to Negative Ratio	RATIO	BAU	159 Children; 6 Teachers; 2 Schools	Low	10 to 14	Disruptive/Off-Task Behaviours	Table 2
1015	Conduct Problems Prevention Research Group (2011)	Fast Track Prevention Intervention	FAST	Workshop	891 Children; 55 Schools	Sufficient	6.5 to 16.5	Conduct problems	Table 2
1091	Kapalka (2006)	Avoiding Repetitions	AVOID	BAU	86 Teachers; Schools; 6 Districts	Not Reported	5 to 10	Children with ADHD	Table 3
1132	Pepler et al. (1995)	Earlscourt Social Skills Group Programme	ESSGP	BAU	74 Children	Low	7 to 12	Children with Aggressive Behaviour	Table 2

Table 2

Intervention Details

ID	Intervention	Intervention Target	Intervention Subgroup Targets	Facilitation Training Required	Certified Facilitator Required	Intervention Facilitator	Mechanisms Used
23	PR	Multisystem Targets	Target Children; Teacher; Peers	Yes	No	Teachers	Behaviour Management Training; Behaviour Tracking; Psychoeducation
27	STARS	Single Target	Target Children	No	Yes	Behaviour Specialist; Teachers	Social Emotional Learning; Behaviour Tracking
43 a	TALI-P	Single Target	Target Children	No	No	Device	Attention Training
43 b	TALI-B	Single Target	Target Children	No	No	Device	Attention Training
44	FIRST	Multisystem Targets	Target Children; Teachers; Parents; Peers	Yes	Yes	Behaviour Coach; Teachers; Parents	Behaviour Management Training; Multisystem Collaboration; Behaviour Tracking
52	EMR	Single Target	Teachers	Yes	No	Trained Facilitator; Teachers	Social Emotional Learning; Psychoeducation
89	MAGIC	Single Target	Parents	No	No	Parents	Behaviour Management Training; Social Emotional Learning; Psychoeducation
92	PLAY	Single Target	Target Children	No	Yes	Counsellor or Social Worker Certified in Adlerian Play Therapy	Social Emotional Learning; Play Therapy

Table 2 (continued)

ID	Intervention	Intervention Target	Intervention Subgroup Targets	Facilitation Training Required	Certified Facilitator Required	Intervention Facilitator	Mechanisms Used
94	ECM	Single Target	Teachers	Yes	No	Teachers	Behaviour Management Training; Social Emotional Learning; Psychoeducation
95	GBG	Multisystem Targets	Target Children; Teacher; Peers	Yes	No	Teachers	Behaviour Management Training
107	MATE	Dual Targets	Target Children; Parents	No	Yes	Certified Programme Facilitator	Behaviour Management Training; Social Emotional Learning; Multisystem Collaboration; Behaviour Tracking
128	DRC+	Multisystem Targets	Target Children; Teachers; Parents	No	No	Teachers; Parents	Behaviour Management Training; Multisystem Collaboration; Behaviour Tracking; Self-Management Skills Training
166	PSYCHED	Dual Targets	Teachers; Parents	Yes	No	Trained Programme Facilitator	Behaviour Management Training; Psychoeducation
171	SELF	Single Target	Parents	No	No	Parents	Behaviour Management Training
205 a	GBGO	Multisystem Targets	Target Children; Teacher; Peers	No	No	Teachers	Behaviour Management Training

Table 2 (continued)

ID	Intervention	Intervention Target	Intervention Subgroup Targets	Facilitation Training Required	Certified Facilitator Required	Intervention Facilitator	Mechanisms Used
205 b	PATHS+	Multisystem Targets	Target Children; Teacher; Peers	No	No	Teachers	Social Emotional Learning; Self-Management Skills Training
208	KEY	Single Target	Teachers	No	No	Teachers	Behaviour Management Training; Social Emotional Learning
232	EEST	Single Target	Target Children	No	Yes	Facilitator Trained in CBT	Executive Function Training; Self-Management Skills Training; Attention Training
257	STAND	Dual Targets	Target Children; Parents	No	Yes	Clinician or Counsellor Certified in Programme	Behaviour Management Training; Multisystem Collaboration; Behaviour Tracking; Self-Management Skills Training
828	SEAT	Dual Targets	Target Children; Peers	No	No	Teachers	Physical Environment Alterations
861	GREET	Single Target	Teachers	No	No	Teachers	Behaviour Management Training
988	RATIO	Single Target	Teachers	No	No	Teachers	Behaviour Management Training

Table 2 (continued)

ID	Intervention	Intervention Target	Intervention Subgroup Targets	Facilitation Training Required	Certified Facilitator Required	Intervention Facilitator	Mechanisms Used
1015	FAST	Multisystem Targets	Target Children; Teachers; Parents; Peers	No	Yes	Behaviour Specialist	Behaviour Management Training; Social Emotional Learning; Multisystem Collaboration; Behaviour Tracking; Psychoeducation; Self-Management Skills Training; Attention Training; Learning Support
1091	AVOID	Single Target	Teachers	No	No	Teachers	Behaviour Management Training
1132	ESSGP	Multisystem Targets	Target Children; Teachers; Parents; Peers	No	Yes	Certified Programme Facilitator	Social Emotional Learning; Multisystem Collaboration

Self-Management Training and Regulation Strategy (STARS)

STARS (Thompson, 2014b) is a self-management and self-monitoring training programme where behavioural coaches collaborate with children to teach social-emotional skills. Small groups of target students take part in nine scripted lessons focusing on problem solving, goal setting, empathy, reframing mistakes as learning opportunities, self-awareness, and response management. Children actively monitor their behavioural progress and engage in feedback and strategy review sessions.

Tali Train (TALI)

TALI (Kirk et al., 2022) is a touchscreen attention training programme featuring gamified, adaptive tasks that target selective attention, attention orienting, sustained attention, and executive attention. Each twenty-minute session includes four attention-training games. The difficulty level automatically adjusts based on each child's progress. Study 43 comprised two separate studies with different comparison groups. TALI, when compared to a Placebo comparator (TALIP), was identified as study 43a. TALI, when compared to a Business-As-Usual comparator (TALIB), was designated as study 43b.

First Step To Success (FIRST)

FIRST (Seeley et al., 2009) is a multisystem early prevention programme involving classroom coaching, home support, and skills coaching for children. All students are briefly monitored to identify those displaying at-risk behaviours. A behaviour coach collaborates with teachers and students in the classroom to implement a points-based reward behaviour management system and gradually withdraws from the classroom. Parents participate in weekly behavioural coaching on positive reinforcement strategies and receive ongoing feedback.

E-M-R Establish, Maintain, Restore (EMR)

EMR (Cook, et al., 2018) is a structured programme designed to develop a common language and practice to enhance teacher-student relationships and students' sense of belonging through three stages: Establish–Maintain–Restore. EMR aims to teach social practices that occur in healthy relationships during phases of closeness and conflict. Teachers are trained in EMR and utilise relationship reflection forms to assess relationships with students and identify areas of need.

1-2-3 Magic (MAGIC)

MAGIC (Chesterfield et al., 2021) is a parent training program that teaches behavioural strategies for discipline, encouraging good behaviour, building relationships, and fostering self-emotion regulation. Parents learn clear and straightforward techniques to use. The program includes operant learning principles, social learning theories, and cognitive behaviour therapy (CBT) techniques. Several delivery options are available, such as group workshops, one-on-one coaching, online modules, or school-hosted parent evenings.

Adlerian Play Therapy (PLAY)

PLAY (Meany-Walen et al., 2014) combines directive and non-directive play techniques, especially beneficial for children with poor social skills, low self-esteem, those who have experienced trauma, or those who have an increased need to control. Play therapy offers a safe environment to practice social skills and explore thoughts and feelings through play. A skills checklist is included to assist counsellors.

Errorless Classroom Management (ECM)

ECM (Klopfer et al., 2019) is a preservice teacher education programme that teaches key behavioural components in the classroom (compliance, focus, social interaction) and behaviour management strategies to promote these. ECM concentrates on preventing errors in identified key behavioural components by teaching scaffolding (modelling and high-probability prompts), providing immediate positive reinforcement, and gradually withdrawing prompts as students become independently successful in these key behaviours.

Good Behaviour Game (GBG)

GBG (Ashworth et al., 2020) is a classroom behaviour management game that involves four clear behavioural rules. During the game, students are grouped into teams, and the teacher monitors team behaviour without intervening, awarding negative scores to teams each time a rule is broken. At the end of the game, positive reinforcement, such as stickers, is given to groups with four or fewer rule infractions.

Mate-Tricks (MATE)

MATE (O'Hare et al., 2015b) is a manualised after-school programme designed for children and parents living in socioeconomically disadvantaged areas. MATE involves various modes of delivery, including child-only, parent-only, and parent-child sessions. Child group sessions focus on social and emotional learning activities based on the Strengthening

Families Programme and the Coping Power Programme. Parent sessions emphasise positive reinforcement and practical communication skills. All sessions are run by trained facilitators.

Daily Report Card and Parent–Teacher Consultation (DRC+)

In the DRC+ (Murray et al., 2008) intervention, parents and teachers meet with the assistance of a consultant to identify target behaviours, set clear and specific goals, and monitor progress every two to three weeks. Consultants facilitate all meetings, organise the necessary resources, and provide support through written summaries of meetings and information about rewards. They also contact parents once to review and troubleshoot the reward system at home. The consultation process is manualised with scripted guidance.

Psychoeducation (PSYCHED)

PSYCHED (Hosseinnia et al., 2024) is delivered in a lecture format by a facilitator trained in ADHD to parents and teachers. It covers characteristics of ADHD, how adult behaviour and what they choose to focus on impacts children with ADHD, and apt behaviour management techniques. The sessions involve role-play and active discussion to support learning. Parent-targeted PSYCHED is based on the Defiant Children parent training manual, while teacher-targeted PSYCHED follows the Classroom Accommodations for Children with ADHD manual.

Self-Help Parenting (SELF)

SELF (de Jong et al., 2023) is a parenting program that includes a chaptered manual and online modules used simultaneously. The programme features educational training modules and tasks to practice techniques. Parents choose three to five target behaviours and track those behaviours alongside their confidence as the program advances.

Good Behaviour Game Only (GBGO)

GBGO (Bradshaw et al., 2020) is the same classroom behaviour management game intervention used in study 95 GBG (Ashworth et al., 2020). Study 205 included two separate interventions compared to the same control. Study 205a involved GBGO compared to a Business-As-Usual comparator.

GBG and Promoting Alternative Thinking Strategies (PATHS+)

PATHS+ (Bradshaw et al., 2020) in the study labelled 205b included GBG and the Promoting Alternative Thinking Strategies (PATHS) programme compared to a Business-As-Usual comparator. PATHS focuses on social-emotional learning and was designed to

complement the behavioural focus of the GBGO. PATHS includes structured lessons that utilise stories, role-play, and discussions to foster skills related to social awareness, self-management, relationship skills, and decision-making. Sessions are held two to three times a week.

Key2Teach (KEY)

KEY (Hoogendijk et al., 2020) is a teacher coaching programme with a social-emotional curriculum designed to develop interpersonal knowledge and skills. A coach initially works with teachers to profile student-teacher relationships. They then focus on teaching teachers skills related to perspective-taking and functional behaviour analysis. During sessions, positive teaching behaviours are identified and summarised in short key phrases. In class, coaches sit at the back, providing feedback to the teacher via an in-ear speaker using the key phrases identified earlier. Lessons are videotaped and reviewed in subsequent sessions.

Ecological Executive Skill Training (EEST)

EEST (Qian et al., 2017b) is a manualised programme where a therapist trains children with ADHD to enhance executive skills using Cognitive Behavioural Therapy (CBT) methods. Groups of six to eight families attend the initial and final weekly sessions, during which parents learn how to support children in completing tasks at home. Children participate in all sessions alongside a therapist.

Supporting Teens' Academic Needs Daily (STAND)

STAND (Sibley et al., 2013) involves teaching parents and children with ADHD how to collaborate to improve academic functioning. STAND is a manualised programme with weekly family sessions over three months that teach and scaffold organisational skills. An optional additional two months of problem-solving skills training are also offered. Monthly parent group sessions are held to develop behaviour management skills needed to implement the STAND programme. After family sessions, parents are coached to arrange a meeting with themselves, the child, and the teacher. The aim of this meeting is to request support at school with academic interventions and to establish a way to monitor progress.

Seating Chart (SEAT)

The SEAT (Braun et al., 2020) intervention seeks to reduce disruptive behaviour among peers. SEAT starts with students ranking their peers based on likability. Pairs where

at least one student expressed disliking the other are seated closer together in the classroom.

Positive Greetings at the Door (GREET)

GREET (Cook, Fiat, et al., 2018) is a three-step preventive classroom management strategy designed to reach all students, rather than targeting specific students. GREET aims to build connections and provide positive precedents for the day ahead. Teachers receive tell-show-do training with feedback to warmly greet students, remind them of behaviour expectations, and encourage engagement in learning as they enter the classroom.

5:1 Positive to Negative Ratio (RATIO)

RATIO (Cook et al., 2017) starts with training teachers in positive reinforcement strategies such as emphasising positive behaviour, providing specific verbal and nonverbal praise and approval. In the classroom, teachers then wear a prompting device (MotivAider®) aimed at increasing the ratio of positive to negative interactions with their students.

Fast Track Prevention Intervention (FAST)

FAST (Conduct Problems Prevention Research Group, 2011) is a long-term, multi-component programme designed for children identified as high-risk of developing conduct disorder. It comprises parent training and home visits, parent-child interactive group social skills training, academic support for children, social-emotional skills training, peer sessions, coordination with progressive classrooms, PATHS, a school transition programme, student coping groups, and parent support during adolescence, along with vocational and life skills training. FAST aims to target different areas influencing child development. Selected interventions are adjusted based on triannual assessment of risk and protective factors.

Avoiding Repetitions (AVOID)

Over three sessions, AVOID (Kapalka, 2006) offers teachers brief psychoeducation on ADHD and training in Barkley's 'Giving Effective Commands' procedure to help break repetitive cycles. This involves giving one clear instruction, waiting, and then applying consequences if the instruction is not followed. AVOID aims to prevent escalating conflict patterns and encourages teachers to be more deliberate and thoughtful about the consequences.

Earlscourt Social Skills Group Programme (ESSGP)

ESSGP (Pepler et al., 1995) is a social learning-based behavioural and social-cognitive training programme conducted at school. Small groups of children attend sessions covering at least 6 core modules. These include listening, anger management, joining a group, following instructions, responses to teasing, avoiding fights, and problem-solving. ESSGP aims to enhance self-control, social skills, and compliance. Parents, teachers, and classroom peers are all offered training in these modules.

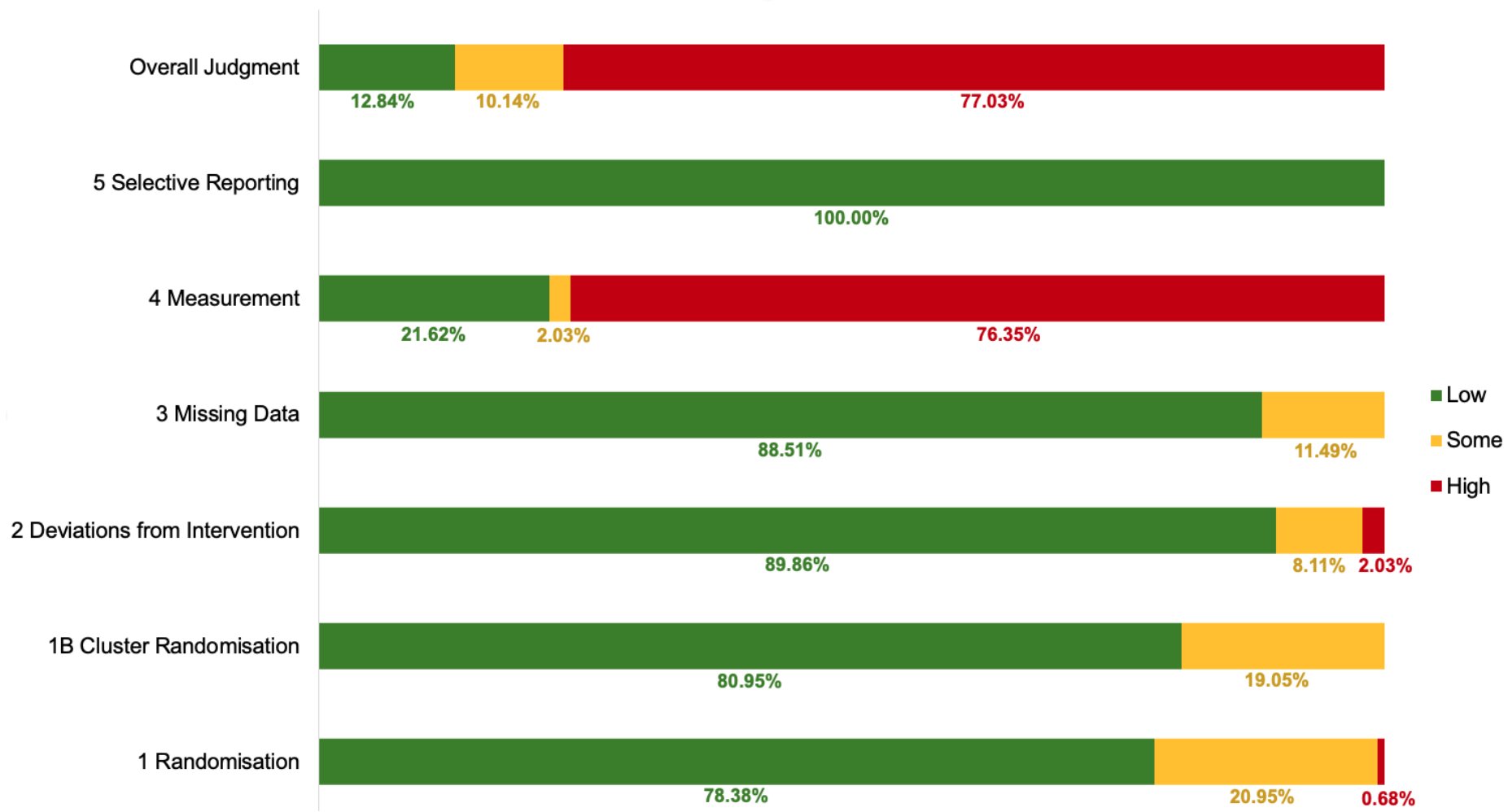
Risk of Bias

The ROB2 assesses the level of *concern about risk* of bias across five domains that reflect the quality of study methods and execution. A ROB judgement was assigned for each included outcome. The level of risk – or systematic error - introduced into studies was rated as low, some, or high concern, which informs the confidence that can be placed in reported effects. ROB judgements help clarify how and where bias might have occurred within each outcome. Low concern indicates that reported effects are likely close to their true values. Some concern suggests that potential systematic error was introduced in at least one area and *could* affect the credibility of the reported effects. High concern means probable systematic error was found in more than one area, or that systematic error likely occurred in at least one area, reducing the accuracy of the reported effects. It is important to remember when interpreting ROB judgments that they indicate the *likelihood* of systematic error being introduced, not whether it actually was (Higgins et al., 2022, p. 179).

The overall ROB was generally rated as high (77.03% high) (see Figure 4). This was influenced especially by domain four relating to the procedures and execution of measurement of outcomes (76.35% high).

Figure 4

Risk of Bias Judgments Forest Plot



Note. This forest plot illustrates the percentage of low, some, and high risk judgments across each ROB2 domain for all included outcomes.

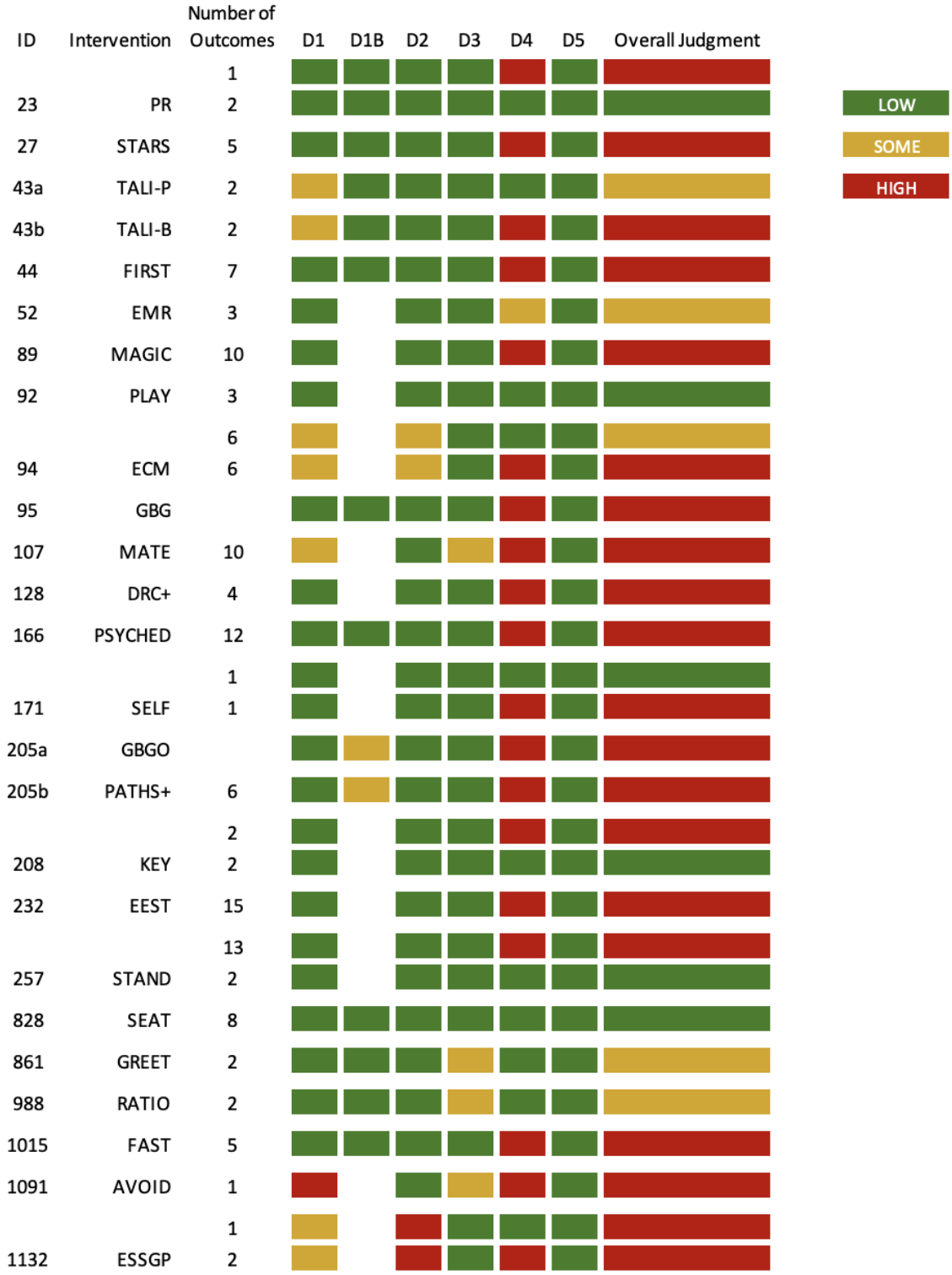
Measurement methods in included studies often involved assessments by unblinded raters, who were frequently also participants. This significantly impacts confidence in reported effects due to the high potential for bias from the information source of those effects. Out of the 148 reported effects, 114 were rated as high overall ROB, 15 as moderate, and 19 as low. Only PLAY and SEAT were rated low across all measured outcomes (Braun et al., 2020; Meany-Walen et al., 2014). All studies demonstrated transparent reporting of data analyses and labelled any post-hoc calculations accordingly.

An overview of the ROB judgments across included studies is shown in a traffic light chart. Outcome measures for the same judgments within a study are grouped and numbered along the side (see Figure 5).

Domain 1A concentrates on the randomisation process and evaluates whether the methods employed might have influenced differences between groups by introducing confounding factors, thereby reducing the ability to attribute effects solely to the intervention. Domain 1B pertains only to cluster-randomised studies and examines the sequence of randomisation in relation to the timing of cluster group allocation. AVOID was the only study assigned a high risk judgment in domain 1, as they did not employ a genuinely random method of assigning participants to groups (Kapalka, 2006). TALI, ECM, MATE, and ESSGP were rated with some concern regarding risk in domain 1. TALI exhibited significant baseline differences in age between groups, although these were reported, and additional results were provided based on these differences (Kirk et al., 2022). ECM participants provided their consent after being informed of their allocation. Significantly more individuals who learned they were assigned to the control group dropped out in comparison to those assigned to the experimental group (Klopfer et al., 2019). Bradshaw et al., (2020) did not report the pretest balance of groups of GBGO or PATHS+ with the comparator condition. The authors studying the MATE intervention did not consider the implications of their randomisation process, which resulted in children from the same families being split across different locations. Siblings within Cohort 1 were reallocated together in the intervention group (O'Hare et al., 2015a). The subsequent cohorts were grouped by families before randomisation. ESSGP did not disclose their randomisation process (Pepler et al., 1995).

Figure 5

ROB2 Judgments Across Studies



Note. D1A = Domain 1A Randomisation Process, D1B = Domain 1B Timing of Randomisation (Cluster Randomised Studies Only), D2 = Domain 2 Deviations from Interventions, D3 = Domain 3 Missing Outcome Data, D4 = Domain 4 Measurement of the Outcome, D5 = Domain 5 Selection of Reported Results, and OJ = Overall Judgment.

Domain 2 evaluates deviations from intended interventions caused by knowledge of allocation, which can either exaggerate or underestimate the actual effects of the intervention. TALIP, EMR, PLAY, GREET, and RATIO completely avoided this risk by employing a placebo or attentional control group to help blind participants to their allocation. ECM was rated with some risk, and ESSGP was rated with high risk. Teachers who consented and remained in the ECM study were more likely to be in the intervention group due to their awareness of allocation. ESSGP also did not provide enough information to thoroughly assess this risk.

Domain 3 assesses whether missing data causes bias by being affected by either the measured outcome or group assignment. Students with certain characteristics might not perform well in an intervention, resulting in a confounding attrition effect where only those who enjoy the intervention stay in the study. However, attrition and missing data are common in research studies. One way to reduce the risk of systematic error from missing data is the proper use of imputation for non-completers. Nonetheless, MATE, GREET, RATIO, and AVOID received some concern of ROB, as none of these studies reported how they handled missing data.

Domain 4 assesses the validity of measurement tools and procedures. Unblinded raters, who were aware of the group allocation of those they were assessing, made up 75.68% of all outcomes. Furthermore, the EMR study did not report whether raters of any of its three outcome measures were blinded (Cook et al., 2018). A large proportion of the effects reported across the included studies probably do not accurately reflect true effects.

Domain 5 assesses whether researchers properly planned their study and followed the original plan rather than covertly using post-hoc methods to highlight stronger effects. In the included studies, authors were transparent about their data analysis methods, and when post hoc analysis was used, it was clearly reported as such and not included in the main results section. Cluster randomised trials provided clear descriptions of how they analysed data at child, class, and school levels.

Non-reporting of information required in the ROB2 assessments occurred in 24% of outcomes. Nearly a quarter of studies had domain risk ratings elevated to some concern due to unknown systematic error. The reporting of study methods and procedures was inadequate, considering the empirical nature of RCTs. One notable study, AVOID, showed an

increased risk of bias across most domains (Kapalka, 2006). Its randomisation process was not genuinely random, measurement instruments were modified without assessing validity, and key details about study methods were omitted, affecting transparency and the ability to make informed judgments. Although I included it in the review, the high risk of systematic error is evident. The reported effects of AVOID should be interpreted with caution. This does not *imply* that the AVOID intervention is invalid or ineffective; rather, it indicates that the study was more likely to yield unreliable results. A table summarising ROB2 judgments, justifications, and bias impacts for each reported outcome has been uploaded to OSF.

Results of Individual Studies

When reporting results by vote counting, the direction of effect for all outcomes must be reported, regardless of whether the effects are significant. This is because vote counting occurs across studies with varying power to detect effects. Reporting the direction of effect for all outcomes promotes fairness. However, reporting all outcomes as beneficial or adverse can risk misinterpreting the effectiveness of interventions when reported effects were not necessarily accurate or related to the intervention. To ensure clarity on the level of confidence in reported results and minimise misinterpretation, Table 3 presents the results from individual studies in sections based on the quality of evidence supporting each outcome (high, moderate, very low). Outcomes are then organised by statistical significance (presented as a tick or cross) and by power (presented as a low or full battery). This clarifies if the reported outcome was attributable to the intervention statistically ($p < .05$) and the study's capacity to detect effects statistically. After safeguarding interpretation, individual results are then sectioned by the direction of effect. This helps identify whether the reported results demonstrated an adverse or beneficial impact.

Types of Outcome Measures

Rating scales were employed to assess most outcomes (80.41%). Teachers and parents were the primary sources for outcome measures, accounting for 39.19% and 31.08% of all measures, respectively. At much lower frequencies, target children (9.46%), peers (6.76%), trained observers (6.76%), and trained coders (4.73%) were involved in outcome assessments. Objective coding, using a checklist of physical materials, represented 1.35% of outcomes measured, while objective devices such as actigraphy monitors were used for one outcome measure (0.68%).

Significance and Size of Effects

Ignoring statistical significance and power, most outcomes were found to have a beneficial effect (85.14%). Half of all outcome effects were statistically attributed to the intervention (50.68%). However, nearly half of the included studies were underpowered to detect effects (45.27%). Significance only indicates whether the effect was attributed to the intervention; when assessing the effectiveness of interventions, effect sizes are crucial for determining whether any effects observed were truly meaningful. An effect may be statistically significant but have only a minor impact compared to the control group. A large proportion of effect sizes reported across outcomes were not disclosed. Reporting of effect sizes for all outcomes was as follows: non-reported 32.43%, trivial 23.65%, large 17.57%, moderate 16.22%, and small 10.14%. Effect sizes are generally not reported for non-significant effects. Nonetheless, the effect size was not reported for 20 of the 75 outcomes that were statistically significant. Of the 73 outcomes that were not statistically significant, the effect size was not reported for 38.36%. My review found that over a quarter of the statistically significant results (26.67%) lacked reported effect sizes.

Certainty of Reported Effects

High-Quality Evidence. SEAT, PLAY, SELF, KEY, STAND, and PR contribute 18 outcomes with high certainty ratings (Braun et al., 2020; de Jong et al., 2023; Hoogendijk et al., 2020; Meany-Walen et al., 2014; Sibley et al., 2013; Veenman et al., 2017). This accounts for 12.16% of all outcomes included in my review (see Table 3). Of these, 10 outcomes identified no effect, indicating the observed effect was not caused by the intervention. Among the eight effects confidently linked to the intervention, SEAT had two trivial to small adverse effects on overt aggression and class-wide cooperation; the others were beneficial. PLAY demonstrated moderate to large beneficial effects on disruptive, externalising, and on-task behaviours, as well as teacher stress. Despite low power, SELF showed small beneficial effects on daily externalising behaviours, and STAND exhibited large beneficial effects on planner use for academic purposes. The remaining results with a high certainty rating confidently indicate that the interventions do not have any particular impact on these outcomes.

Moderate Quality Evidence. TALIP, EMR, ECM, GREET, RATIO, and STAND contribute 15 outcomes with moderate certainty ratings (Cook, et al., 2018; Cook et al., 2017; Cook,

Fiat, et al., 2018; Kirk et al., 2022; Klopfer et al., 2019; Sibley et al., 2013). That accounts for 10.14% of all included outcomes (see Table 3). Of these, STAND showed moderate beneficial effects on children's organisation of school materials that were *not link*ed to the intervention. However, the study had low power to detect effects – if the sample had been larger, these effects might have been attributed to the intervention. The remaining results in Table 4 demonstrate small to large beneficial effects across these outcomes, despite the studies all having low power to detect effects. Nonetheless, there is only moderate certainty that these effects are valid, as concerns about bias were present in the study procedures used to obtain these results.

Very Low Quality Evidence. Out of all 148 outcomes, 115 exhibited very low certainty regarding the validity of the reported effects. That is 77.7% of all reported outcomes in my review (see Table 3). All but one outcome rated with very low certainty were due to studies measuring the outcomes with unblinded raters. The other one, ESSGP, was rated as such due to a compounding lack of reporting of procedural details. Studies with 100% very low certainty ratings were STARS, TALIB, FIRST, MAGIC, GBG, MATE, DRC+, PSYCHED, GBGO, PATHS+, EEST, FAST, and AVOID (Ashworth et al., 2020; Bradshaw et al., 2020; Chesterfield et al., 2021; Conduct Problems Prevention Research Group, 2011; Hosseinnia et al., 2024; Kapalka, 2006; H. E. Kirk et al., 2022; Murray et al., 2008; O'Hare et al., 2015b; Qian et al., 2017b; Seeley et al., 2009; Thompson, 2014b). These studies employed unblinded raters for all reported outcomes, increasing the risk of detection bias and potentially overstating intervention effects (Hróbjartsson et al., 2013). Over half of the included studies (56.52%) demonstrated very low certainty in reported effects. While results with very low certainty might still reflect the true effect of the intervention, poor study rigour means most findings should be interpreted cautiously. In this section, while power and p-value cutoffs can suggest the real effects of interventions, these are overshadowed by the lack of certainty stemming from insufficient study procedures and execution.

Table 3

Results of Individual Studies

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
High Quality of Evidence						
High certainty that the effect is attributable to the intervention.						
Adverse						
			Overt Aggression	% Nominated	Peer-Rated	Trivial
	828	SEAT	Class-Wide Cooperation	Scales	Peer-Rated	Small
Beneficial						
			Externalising Behaviour	Scales	Teacher-Rated	Moderate
			Teacher Stress from the Teacher-Child Relationship	Scales	Teacher-Rated	Moderate
			Disruptive Behaviours	Observation	Trained Observer	Large
	92	PLAY	On-Task Behaviours	Observation	Trained Observer	Large
High certainty the effect is attributable to the intervention despite low power.						
Beneficial						
	171	SELF	Daily Externalising Behaviour	Scales	Trained Coder	Small
	257	STAND	Planner Use (Academic Engagement/Organisation)	Coded Materials	Objective Coding	Large
High certainty that the intervention did not change the outcome.						
Adverse						
	23	PR	Peer Acceptance	Scales	Peer-Rated	NR
	208	KEY	Teacher Interaction Skills Emotional Support	Observation	Trained Observer	NR
			Class-Wide Cohesion	Scales	Peer-Rated	Trivial
			Class-Wide Conflict	Scales	Peer-Rated	Trivial
			Class-Wide Isolation	Scales	Peer-Rated	Trivial
	828	SEAT	Prosocial Behaviour	% Nominated	Peer-Rated	Trivial

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
High Quality of Evidence						
High certainty that the intervention did not change the outcome.						
Beneficial						
	23	PR	Hyperactivity	Actigraphy Monitor	Objective Device	NR
	208	KEY	Teacher Interaction Skills Proximity	Scales	Child-Rated	Trivial
			Acceptance	% Nominated	Peer-Rated	Trivial
	828	SEAT	Rejection	% Nominated	Peer-Rated	Trivial
Moderate Quality of Evidence						
The effect is attributable to the intervention despite low power. Though, poor study procedures lowers certainty of the validity of effects.						
Beneficial						
	43a	TALIP	Hyperactivity/Inattention (ADHD Combined)	Scales	Teacher-Rated	NR
			Disruptive Behaviour	Observation	Trained Observer	Small
			Student-Teacher Relationship Quality	Scales	Teacher-Rated	Moderate
	52	EMR	Academic Engaged Time	Observation	Trained Observer	Large
			Inadequate Teaching Strategies in Compliance Scenario	Videoed Observation	Trained Coders	NR
			Inadequate Teaching Strategies in Off-Task Scenario	Videoed Observation	Trained Coders	Moderate
			Proactive Teaching Strategies in Compliance Scenario	Videoed Observation	Trained Coders	Large
			Proactive Teaching Strategies in Off-Task Scenario	Videoed Observation	Trained Coders	Large
	94	ECM	Reactive Teaching Strategies in Compliance Scenario	Videoed Observation	Trained Coders	Large
			Academic Engaged Time	Observation	Trained Observer	Large
	861	GREET	Disruptive Behaviour	Observation	Trained Observer	Large
			Academic Engaged Time	Observation	Trained Observer	Large
	988	RATIO	Disruptive Behaviour	Observation	Trained Observer	Large

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
Moderate Quality of Evidence						
A valid effect caused by the intervention was *not* found. Though, power was insufficient to detect effects.						
Beneficial						
	257	STAND	Organisation (of School Materials)	Coded Materials	Objective Coding	Moderate
Very Low Quality of Evidence						
The effect is attributable to the intervention. Though, study procedures seriously lowered certainty in the validity of effects.						
Adverse						
			Antisocial Behaviour PSBQ	Scales	Child-Rated	Trivial
			Liberal Parenting	Scales	Child-Rated	Trivial
	107	MATE	Authoritarian Parenting	Scales	Child-Rated	Small
Beneficial						
			Student-Teacher Relationship	Scales	Teacher-Rated	Small
			Authority Defiance	Scales	Teacher-Rated	Moderate
			Disruptive Behaviour	Scales	Teacher-Rated	Moderate
	27	STARS	Social Competence	Scales	Teacher-Rated	Moderate
			ADHD-Related Defiance/Aggression	Scales	Parent-Rated	Moderate
			ADHD-Related Executive Function	Scales	Parent-Rated	Moderate
			ADHD-Related Hyperactivity/Impulsivity	Scales	Parent-Rated	Moderate
			Parental Distress	Scales	Parent-Rated	Moderate
			Difficult Child	Scales	Parent-Rated	Large
			Intensity of Child Behaviour	Scales	Parent-Rated	Large
			Parent-Child Dysfunctional Interaction	Scales	Parent-Rated	Large
	89	MAGIC	Problem Behaviours	Scales	Parent-Rated	Large

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
Very Low Quality of Evidence						
The effect is attributable to the intervention. Though, study procedures greatly lowered certainty in the validity of effects.						
Beneficial						
	107	MATE	Antisocial Behaviour CBCL	Scales	Teacher-Rated	Trivial
	205b	PATHS+	Social Competence	Scales	Teacher-Rated	Trivial
			Student–Teacher Conflict	Scales	Teacher-Rated	Small
	208	KEY	Student–Teacher Closeness	Scales	Teacher-Rated	Moderate
			Hyperactivity-Impulsivity	Scales	Parent-Rated	NR
			Initiation Executive Function Daily Behaviours	Scales	Parent-Rated	NR
			Learning and School Impairment from Behavioural Difficulties	Scales	Parent-Rated	NR
	232	EEST	Life Skill Impairment (Managing and Organising Self)	Scales	Parent-Rated	NR
The effect is attributable to the intervention despite low power. However, study procedures greatly lowered certainty in the validity of effects.						
Beneficial						
	43b	TALIB	Hyperactivity/Inattention (ADHD Combined)	Scales	Teacher-Rated	NR
			ODD Symptoms	Scales	Teacher-Rated	Moderate
			Adaptive Behaviour	Scales	Teacher-Rated	Large
			Hyperactivity	Scales	Teacher-Rated	Large
			Inattention	Scales	Teacher-Rated	Large
			Maladaptive Behaviour	Scales	Teacher-Rated	Large
	44	FIRST	Social Skills	Scales	Teacher-Rated	Large

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
Very Low Quality of Evidence						
The effect is attributable to the intervention despite low power. Though, study procedures greatly lowered certainty in the validity of effects.						
Beneficial						
			Teacher Attitudes - Positive Emotions Towards Children with Challenging Behaviour	Scales	Teacher-Rated	Moderate
			Teaching Style - Psychological Pressure (Enragement and Disappointment)	Scales	Teacher-Rated	Moderate
	94	ECM	Teacher Attitudes - Positive Reactions Towards Children with Challenging Behaviour	Scales	Teacher-Rated	Large
	128	DRC+	Academic Productivity	Scales	Teacher-Rated	Moderate
			ADHD Combined	Scales	Parent-Rated	NR
			ADHD Combined	Scales	Teacher-Rated	NR
			Hyperactive-Impulsive	Scales	Teacher-Rated	NR
			Hyperactive-Impulsive	Scarles	Parent-Rated	NR
			Inattentive	Scales	Teacher-Rated	NR
			Inattentive	Scales	Parent-Rated	NR
			Parent Attitude	Scales	Parent-Rated	NR
			Parent Behaviour	Scales	Parent-Rated	NR
			Parent Knowledge	Scales	Parent-Rated	NR
			Teacher Attitude	Scales	Teacher-Rated	NR
			Teacher Behaviour	Scales	Teacher-Rated	NR
	166	PSYCHED	Teacher Knowledge	Scales	Teacher-Rated	NR
	171	SELF	Externalising Behaviour	Scales	Parent-Rated	Moderate

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
Very Low Quality of Evidence						
The effect is attributable to the intervention despite low power. Though, study procedures greatly lowered certainty in the validity of effects.						
Beneficial						
			ADHD-Related Academic Problems	Scales	Teacher-Rated	Trivial
			Conflict with Parent	Scales	Child-Rated	Moderate
			DBD Symptom Hyperactivity/Impulsivity Severity	Scales	Teacher-Rated	Moderate
			DBD Symptom Inattention Severity	Scales	Parent-Rated	Large
	257	STAND	DBD Symptom ODD Severity	Scales	Parent-Rated	Large
	1091	AVOID	Non-Compliance	Scales	Teacher-Rated	NR
	1132	ESSGP	Externalising Behaviour	Scales	Teacher-Rated	Moderate
The effect is not attributable to the intervention. Though, study procedures greatly lowered certainty in the validity of effects.						
Adverse						
			Disruptive Behaviour	Scales	Teacher-Rated	Trivial
			Prosocial Behaviour	Scales	Teacher-Rated	Trivial
	95	GBG	Prosocial Behaviour PSBQ	Scales	Child-Rated	Trivial
			Trait Emotional Intelligence	Scales	Child-Rated	Trivial
			Trait Emotional Intelligence	Scales	Teacher-Rated	Trivial
	107	MATE	Antisocial Behaviour CBCL	Scales	Child-Rated	Small
			Academic Engagement	Scales	Teacher-Rated	Trivial
			Aggressive-Disruptive	Scales	Teacher-Rated	Trivial
			Hyperactivity	Scales	Teacher-Rated	Trivial
	205a	GBGO	Peer Relations	Scales	Teacher-Rated	Trivial
Beneficial						
	23	PR	ADHD Symptoms	Observation	Trained Observer	NR
	27	STARS	Student-Teacher Relationship	Scales	Child-Rated	Trivial

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
Very Low Quality of Evidence						
The effect is not attributable to the intervention. Though, study procedures greatly lowered certainty in the validity of effects.						
Beneficial						
			ADHD-Related Peer Relations	Scales	Parent-Rated	Small
	89	MAGIC	ADHD-Related Inattention	Scales	Parent-Rated	Moderate
	95	GBG	Concentration Problems	Scales	Teacher-Rated	Trivial
			ADHD-Related Behaviours	Scales	Teacher-Rated	Trivial
			Paternal Relationship	Scales	Child-Rated	Trivial
			Supportive Parenting	Scales	Child-Rated	Trivial
	107	MATE	Maternal Relationship	Scales	Child-Rated	Small
			Emotion Regulation	Scales	Teacher-Rated	Trivial
	205a	GBGO	Social Competence	Scales	Teacher-Rated	Trivial
			Academic Engagement	Scales	Teacher-Rated	Trivial
			Aggressive-Disruptive	Scales	Teacher-Rated	Trivial
			Emotion Regulation	Scales	Teacher-Rated	Trivial
			Hyperactivity	Scales	Teacher-Rated	Trivial
	205b	PATHS+	Peer Relations	Scales	Teacher-Rated	Trivial
			Emotion Control Executive Function Daily Behaviours	Scales	Parent-Rated	NR
			Family Functioning Impairment	Scales	Parent-Rated	NR
			Inattention	Scales	Parent-Rated	NR
			Inhibition Executive Function Daily Behaviours	Scales	Parent-Rated	NR
			Monitoring Executive Function Daily Behaviours	Scales	Parent-Rated	NR
			Organizing Executive Function Daily Behaviours	Scales	Parent-Rated	NR
			Planning Executive Function Daily Behaviours	Scales	Parent-Rated	NR
			Risky Activities (Impulsivity and High-Risk Behaviours)	Scales	Parent-Rated	NR
			Shifting Executive Function Daily Behaviours	Scales	Parent-Rated	NR
			Social Activities Impairment (Participation and Functioning)	Scales	Parent-Rated	NR
	232	EEST	Working Memory Executive Function Daily Behaviours	Scales	Parent-Rated	NR

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
Very Low Quality of Evidence						
The effect is not attributable to the intervention. Though, study procedures greatly lowered certainty in the validity of effects.						
Beneficial						
			Risk Development ADHD	Coded Interview	Parent-Rated	NR
			Risk Development CD	Coded Interview	Child-Rated	NR
			Risk Development CD	Coded Interview	Parent-Rated	NR
			Risk Development ODD	Coded Interview	Child-Rated	NR
	1015	FAST	Risk Development ODD	Coded Interview	Parent-Rated	NR
The effect is not attributable to the intervention. Though, power was insufficient to detect effects and study procedures greatly lowered certainty in the validity of effects.						
Adverse						
	43a	TALI-P	Hyperactivity/Inattention (ADHD Combined)	Scales	Parent-Rated	NR
Beneficial						
	43b	TALIB	Hyperactivity/Inattention (ADHD Combined)	Scales	Parent-Rated	NR
	44	FIRST	Social Skills Home-Based	Scales	Parent-Rated	Moderate
			Reactive Teaching Strategies in Off-Task Scenario	Videoed Observation	Trained Coders	NR
			Teacher Attitudes - Negative Beliefs Towards Children with Challenging Behaviour	Scales	Teacher-Rated	NR
			Teacher Attitudes - Positive Planned Behaviours Towards Children with Challenging Behaviour	Scales	Teacher-Rated	NR
	94	ECM	Teaching Style - Warmth & Support	Scales	Teacher-Rated	NR
			Attention Impairment	Scales	Teacher-Rated	Trivial
			Academic Impulse Control	Scales	Teacher-Rated	Small
	128	DRC+	Deportment Impairment (Rule-Inappropriate Social-Behavioural Impairment)	Scales	Teacher-Rated	Small
			Academic Functional Impairment	Scales	Teacher-Rated	Small
	257	STAND	Caregiver Strain	Scales	Parent-Rated	Small

Table 3 (continued)

Effect Direction	ID	Intvn.	Outcome	How Measured	Rater	Effect Size
Very Low Quality of Evidence						
⊗ **The effect is not attributable to the intervention. Though, power was insufficient to detect effects and study procedures greatly lowered certainty in the validity of effects.**						
Beneficial						
			DBD Symptom Inattention Severity	Scales	Teacher-Rated	Small
			DBD Symptom ODD Severity	Scales	Teacher-Rated	Small
			Academic Functional Impairment	Scales	Parent-Rated	Moderate
			ADHD-Related Academic Problems	Scales	Parent-Rated	Large
			Conflict with Child	Scales	Parent-Rated	Large
			DBD Symptom Hyperactivity/Impulsivity Severity	Scales	Parent-Rated	Large
			Aggression	% Nominated	Peer-Rated	NR
	1132	ESSGP	Externalising Behaviour	Scales	Parent-Rated	NR

Note. Intvn. = Intervention, ■ = Sufficient Power, ▯ = Low Power, ✓ = Effect was statistically significant at $p < .05$, ⊗ = Effect was not statistically significant at $p < .05$.

Synthesis Groupings

Outcomes were categorised into construct domains and subcategories to synthesise similar outcomes to easily review the types of intervention effects reported across included studies. For reviewing moderator effects, outcomes were also assigned to the transactional categories – behavioural and relational domains. Outcomes and studies contributing to the result synthesis are presented in Table 4. The table is divided into behavioural and relational groupings, and within these, outcomes are further grouped by their construct domain and subcategories. A complete list of how outcomes were grouped is in Appendix C. On average, each construct subcategory included 18.49% high certainty ratings from contributing outcomes. Notably, there were no high certainty ratings contributing to the following construct subcategories: Engagement and Organisation Deficits, Attentional Dysregulation, Executive Dysfunction, Emotional Awareness and Regulation, Defiance, Parent Factors, Parent-Child Relationship, and Teacher-Child Relationship.

Characteristics of Groupings

The behavioural domain accounted for 67.57% of all outcomes, which measured outcomes specific to the target children. The relational domain contributed only 32.43% of outcomes, assessing the quality of relationships or including measures of others who influence a child's developmental pathway. Subcategories within the behavioural domain tracked positive outcomes aimed at improvement (Improve) and negative outcomes aimed at reduction (Reduce). Overall, there was a disproportionate focus on measuring negative outcomes, with 70% of behavioural domain outcomes and 47.30% of all outcomes classified as Reduce. Relational domain subcategories include direct measures of relationship quality between target children and others within their ecological system (Direct), and indirect measures of others' impact on factors affecting their relational or transactional quality (Indirect). This grouping was deliberate, recognising the significant influence of relationships and others on a child's developmental trajectory towards externalising disorders. Within the relational domain, indirect measures nearly doubled direct measures of relationship quality, at 62.50% and 37.50%, respectively. Overall, indirect measures made up 20.27% of all outcomes, while direct measures accounted for 12.16%. Outcomes were grouped into five construct domains. Following SWiM guidelines, these constructs were divided into areas of functioning: Academic, Neurocognitive, and Social. Outcomes targeting

relationships did not fit into these construct domains but aligned with direct Relationship Quality and indirect Transactional Quality relational concepts.

Academic Functioning is divided into the subcategories: Active Engagement and Organisational Skills (Improve), Class Disruptive Behaviours (Reduce), and Engagement and Organisation Deficits (Reduce). Active engagement and organisational skills assess not only a child's ability to stay on task but also their level of intentional engagement and the organisational skills that support academic involvement. Class disruptive behaviours measure classroom-based disruptions and off-task behaviours. Engagement and organisation deficits evaluate behaviours and abilities that negatively affect the child's capacity to engage in learning or academic tasks.

Neurocognitive functioning was divided into the following subcategories: Disinhibition (Reduce), Executive Dysfunction (Reduce), and Emotional Awareness and Regulation (Improve). These subcategories all represent skills that are heavily reliant on the proper functioning of brain networks responsible for managing attention, behaviour, and emotions. Attentional dysregulation reflected issues with controlling attention, often referred to as inattention and less often as concentration problems. Disinhibition indicated a child's difficulty inhibiting behaviours, commonly described as impulsivity and hyperactivity across different outcomes. Executive Dysfunction, regarded as the conductor that oversees and coordinates brain networks, measures a child's overall ability to manage different neural systems. This included general or combined measures of ADHD, which reflect a child's capacity to regulate attention and behaviour. Emotional Awareness and Regulation assess a child's ability to understand and manage their emotions, closely linked to the prefrontal cortex and the limbic system. These outcomes are included because emotional intelligence involves not just recognising emotions but also depends on the development of brain areas and networks influencing emotion.

Social Functioning was divided into the following subcategories: Social Competence (Improve), Antisocial Aggressive Behaviour (Reduce), Defiance (Reduce), and Externalising Behaviours Composite (Reduce). Social Competence assesses outcomes related to prosocial behaviours and social interaction skills. Antisocial Aggressive Behaviour measures outcomes related to aggressive behaviours and symptoms linked to CD. Defiance assesses outcomes associated with intentional non-compliance, defiance, and symptoms related to ODD.

Externalising Behaviours Composite covers generalised outcomes concerning externalising or problematic behaviours.

The Transactional Quality construct domain was divided into parent, peer, and teacher factors – representing outcomes that measure the stress, attitude, knowledge, and skills of those nearby to the target children. These factors could influence how they interact with the target child.

The Relationship Quality construct domain was divided into the target children's relationships with parents, peers, and teachers. It reflected outcomes that assessed the conflict, closeness, or the functioning and dysfunction of the child's relationships with each of these groups.

Result Synthesis

The main focus of my review is the effectiveness of interventions for primary-aged children exhibiting externalising behaviours. This is shown through vote counting of the effect direction, with study power, certainty rating, and statistical significance noted for each result. Overall, the quality of research in this area clearly affects the ability to accurately identify intervention effectiveness. Between only one and four outcomes from each construct domain have a high certainty rating, some of which come from studies with limited power. Trends in effect direction suggest that interventions are generally beneficial. However, the validity of these results is highly uncertain and should be interpreted with caution. Statistically significant outcomes made up 52.8% of the results, and about half of these significant effects came from studies with low power. Additionally, 9.38% of outcomes that showed no significant effect are marked as needing cautious interpretation because they come from studies with low power to detect effects.

Within academic functioning, intervention effectiveness for enhancing active engagement and organisational skills, as well as reducing disruptive behaviours in class, mainly showed large beneficial effects with varying data quality (Figure 6).

Table 4

Construct Grouping of Results Included in Synthesis by Behavioural and Relational Domains

ID	Intvn.	Outcome	Rater	Outcome Certainty Rating
Behavioural				
Academic Functioning				
Active Engagement and Organisational Skills (Improve)				
52	EMR	Academic Engaged Time	Trained Observer	Moderate
92	PLAY	On-Task Behaviours	Trained Observer	High
128	DRC+	Academic Productivity	Teacher-Rated	Very Low
205a	GBGO	Academic Engagement	Teacher-Rated	Very Low
205b	PATHS+	Academic Engagement	Teacher-Rated	Very Low
232	EEST	Initiation Executive Function Daily Behaviours	Parent-Rated	Very Low
257	STAND	Planner Use (Academic Engagement/Organisation)	Objective Coding	High
861	GREET	Academic Engaged Time	Trained Observer	Moderate
988	RATIO	Academic Engaged Time	Trained Observer	Moderate
Class Disruptive Behaviours (Reduce)				
27	STARS	Disruptive Behaviour	Teacher-Rated	Very Low
44	FIRST	Maladaptive Behaviour	Teacher-Rated	Very Low
52	EMR	Disruptive Behaviour	Trained Observer	Moderate
92	PLAY	Disruptive Behaviours	Trained Observer	High
95	GBG	Disruptive Behaviour	Teacher-Rated	Very Low
128	DRC+	Deportment Impairment (Rule-Inappropriate Social-Behavioural Impairment)	Teacher-Rated	Very Low
861	GREET	Disruptive Behaviour	Trained Observer	Moderate
988	RATIO	Disruptive Behaviour	Trained Observer	Moderate
Engagement and Organisation Deficits (Reduce)				
89	MAGIC	ADHD-Related Executive Function	Parent-Rated	Very Low
232	EEST	Learning and School Impairment from Behavioural Difficulties	Parent-Rated	Very Low
257	STAND	Academic Functional Impairment	Teacher-Rated	Very Low
Neurocognitive Functioning				
Attentional Dysregulation (Reduce)				
44	FIRST	Inattention	Teacher-Rated	Very Low
89	MAGIC	ADHD-Related Inattention	Parent-Rated	Very Low
95	GBG	Concentration Problems	Teacher-Rated	Very Low
128	DRC+	Attention Impairment	Teacher-Rated	Very Low
166	PSYCHED	Inattentive	Teacher-Rated	Very Low
232	EEST	Inattention	Parent-Rated	Very Low
257	STAND	DBD Symptom Inattention Severity	Teacher-Rated	Very Low

Table 4 (continued)

ID	Intvn.	Outcome	Rater	Outcome Certainty Rating
Behavioural				
Neurocognitive Functioning				
Disinhibition (Reduce)				
23	PR	Hyperactivity	Objective Device	High
44	FIRST	Hyperactivity	Teacher-Rated	Very Low
89	MAGIC	ADHD-Related Hyperactivity/Impulsivity	Parent-Rated	Very Low
128	DRC+	Academic Impulse Control	Teacher-Rated	Very Low
166	PSYCHED	Hyperactive-Impulsive	Teacher-Rated	Very Low
205a	GBGO	Hyperactivity	Teacher-Rated	Very Low
205b	PATHS+	Hyperactivity	Teacher-Rated	Very Low
232	EEST	Hyperactivity-Impulsivity	Parent-Rated	Very Low
257	STAND	DBD Symptom Hyperactivity/Impulsivity Severity	Teacher-Rated	Very Low
Executive Dysfunction (Reduce)				
23	PR	ADHD Symptoms	Trained Observer	Very Low
43a	TALIP	Hyperactivity/Inattention (ADHD Combined)	Teacher-Rated	Moderate
43b	TALIB	Hyperactivity/Inattention (ADHD Combined)	Teacher-Rated	Very Low
107	MATE	ADHD-Related Behaviours	Teacher-Rated	Very Low
166	PSYCHED	ADHD Combined	Teacher-Rated	Very Low
1015	FAST	Risk Development ADHD	Parent-Rated	Very Low
Emotional Awareness and Regulation (Improve)				
107	MATE	Trait Emotional Intelligence	Child-Rated	Very Low
205a	GBGO	Emotion Regulation	Teacher-Rated	Very Low
205b	PATHS+	Emotion Regulation	Teacher-Rated	Very Low
232	EEST	Emotion Control Executive Function Daily Behaviours	Parent-Rated	Very Low
Social Functioning				
Social Competence (Improve)				
27	STARS	Social Competence	Teacher-Rated	Very Low
44	FIRST	Social Skills	Teacher-Rated	Very Low
95	GBG	Prosocial Behaviour	Teacher-Rated	Very Low
107	MATE	Prosocial Behaviour PSBQ	Child-Rated	Very Low
205a	GBGO	Social Competence	Teacher-Rated	Very Low
205b	PATHS+	Social Competence	Teacher-Rated	Very Low
828	SEAT	Prosocial Behaviour	Peer-Rated	High

Table 4 (continued)

ID	Intvn.	Outcome	Rater	Outcome Certainty Rating
Behavioural				
Social Functioning				
Antisocial Aggressive Behaviour (Reduce)				
89	MAGIC	ADHD-Related Defiance/Aggression	Parent-Rated	Very Low
107	MATE	Antisocial Behaviour CBCL	Teacher-Rated	Very Low
205a	GBGO	Aggressive-Disruptive	Teacher-Rated	Very Low
205b	PATHS+	Aggressive-Disruptive	Teacher-Rated	Very Low
828	SEAT	Overt Aggression	Peer-Rated	High
1015	FAST	Risk Development CD	Parent-Rated	Very Low
1132	ESSGP	Aggression	Peer-Rated	Very Low
Defiance (Reduce)				
27	STARS	Authority Defiance	Teacher-Rated	Very Low
44	FIRST	ODD Symptoms	Teacher-Rated	Very Low
257	STAND	DBD Symptom ODD Severity	Teacher-Rated	Very Low
257	STAND	DBD Symptom ODD Severity	Parent-Rated	Very Low
1015	FAST	Risk Development ODD	Parent-Rated	Very Low
1091	AVOID	Non-Compliance	Teacher-Rated	Very Low
Externalising Behaviours Composite (Reduce)				
89	MAGIC	Intensity of Child Behaviour	Parent-Rated	Very Low
92	PLAY	Externalising Behaviour	Teacher-Rated	High
171	SELF	Daily Externalising Behaviour	Trained Coder	High
1132	ESSGP	Externalising Behaviour	Teacher-Rated	Very Low
Relational				
Transactional Quality				
Parent Factors (Indirect)				
89	MAGIC	Parental Distress	Parent-Rated	Very Low
107	MATE	Supportive Parenting	Child-Rated	Very Low
166	PSYCHED	Parent Attitude	Parent-Rated	Very Low
257	STAND	Caregiver Strain	Parent-Rated	Very Low
Peer Factors (Indirect)				
828	SEAT	Class-Wide Cooperation	Peer-Rated	High
Teacher Factors (Indirect)				
92	PLAY	Teacher Stress from the Teacher-Child Relationship	Teacher-Rated	High
94	ECM	Proactive Teaching Strategies in Off-Task Scenario	Trained Coders	Moderate
166	PSYCHED	Teacher Attitude	Teacher-Rated	Very Low
208	KEY	Teacher Interaction Skills Emotional Support	Trained Observer	High

Table 4 (continued)

ID	Intvn.	Outcome	Rater	Outcome Certainty Rating
Relational				
Relationship Quality				
Parent-Child Relationship (Direct)				
89	MAGIC	Parent-Child Dysfunctional Interaction	Parent-Rated	Very Low
107	MATE	Paternal Relationship	Child-Rated	Very Low
232	EEST	Family Functioning Impairment	Parent-Rated	Very Low
257	STAND	Conflict with Parent	Child-Rated	Very Low
Peer-Child Relationship (Direct)				
23	PR	Peer Acceptance	Peer-Rated	High
89	MAGIC	ADHD-Related Peer Relations	Parent-Rated	Very Low
205a	GBGO	Peer Relations	Teacher-Rated	Very Low
205b	PATHS+	Peer Relations	Teacher-Rated	Very Low
232	EEST	Social Activities Impairment (Impaired Social Participation and Functioning)	Parent-Rated	Very Low
828	SEAT	Acceptance	Peer-Rated	High
Teacher-Child Relationship (Direct)				
27	STARS	Student-Teacher Relationship	Child-Rated	Very Low
52	EMR	Student-Teacher Relationship Quality	Teacher-Rated	Moderate
208	KEY	Student–Teacher Closeness	Teacher-Rated	Very Low

Note. Intvn. = Intervention. Outcome groupings are first split by behavioural and relational domains, then by SWiM Construct domain groups, and further into the construct subcategories. In parentheses beside the construct subcategories are the behavioural and relational domain subcategories. Certainty ratings relate to the individual outcomes contributing to each domain

Active engagement and organisational skills have a relatively high level of certainty compared to other construct subcategories, with fewer than half of the contributing outcomes rated as having very low certainty. It is one of only three construct subcategories across all synthesised results where less than half of its contributing outcomes had very low certainty ratings, indicating the generally poor quality of data in this review. In both of these subcategories, the outcomes of the good behaviour game from two different studies were inconsistent with those of other interventions, showing only a trivial adverse effect. The interventions aimed at reducing engagement and organisational deficits demonstrated small to moderate beneficial effects, although with very low certainty across all reported results.

What is notable about the effects within this construct domain is that a substantial proportion (45%) of outcomes showed beneficial effects, despite the low statistical power.

Regarding data quality, measures of neurocognitive functioning, which are strongly linked to measures of ADHD, exhibited the least certainty among all construct domains. Only one beneficial outcome received a high certainty rating in this domain (3.85%), and the effect size for this outcome was not reported (Figure 7). Interventions aimed at reducing attentional dysregulation, disinhibition, and executive dysregulation, while enhancing emotional awareness and regulation, were generally effective. Only the MATE and GBGO interventions showed trivial adverse effects within this construct domain. No measures of emotional awareness and regulation were statistically significant.

The effectiveness within the social functioning construct was again mostly beneficial (Figure 8). With the addition of the SEAT intervention, the MATE, GBG, and GBGO interventions were again inconsistent with other interventions, showing an adverse effect. Intervention effectiveness for reducing antisocial aggressive behaviours was poor. Only one outcome from the MATE intervention showed a statistically significant beneficial effect, indicating the effect could be attributed to the intervention. However, even this effect was trivial in size; it did not reach the threshold for a small effect. Furthermore, the SEAT intervention had an adverse effect on antisocial and aggressive behaviours. Intervention effectiveness in reducing defiant behaviours mostly demonstrated moderately sized significant beneficial effects, although all contributing outcomes for defiant behaviours are rated with very low certainty. Composite outcome measures for externalising behaviours showed small to large beneficial effects, with half of the contributing outcomes rated with high certainty. Social competence effects were inconsistent, but all statistically significant effects showed beneficial outcomes. Notably, despite low power, the FIRST intervention demonstrated large beneficial effects on improving social competence.

Turning to relationally-based measures, the constructs of relationship quality and transactional quality assessed direct and indirect measures of target children's relationships, respectively. Within the relationship quality construct, intervention effectiveness was generally beneficial, except for non-significant adverse impacts on target children's relationships with peers from the GBGO and PR interventions (Figure 9). Although neither of these adverse effects was statistically significant, both stemmed from studies with

sufficient power to detect effects. In fact, concerning peer relationships, all outcomes had sufficient power to identify effects, and no significant effects were observed in this subcategory.

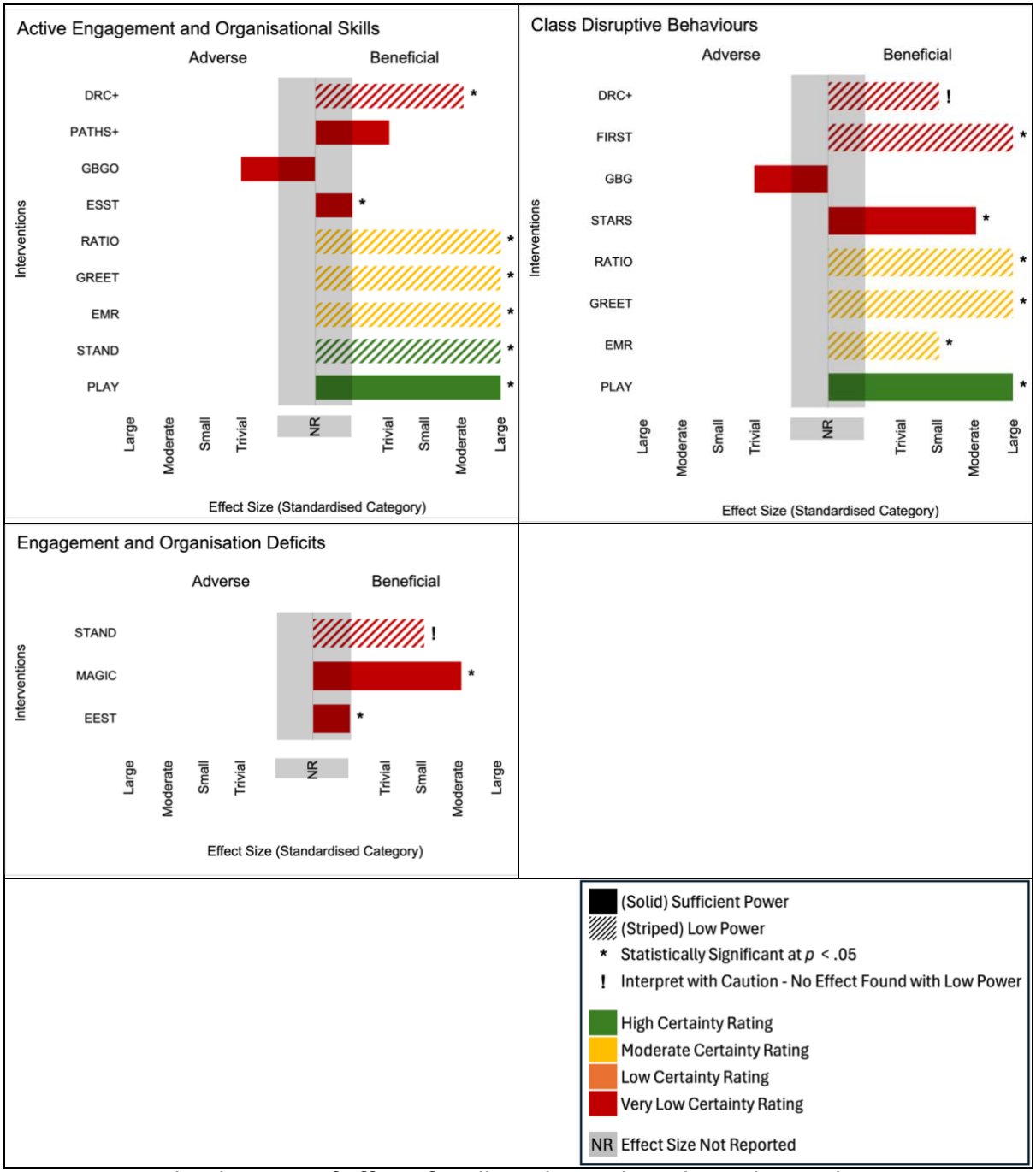
All measures of parent-child relationships were beneficial, although only two were statistically significant, and all had very low certainty ratings. The confidence in the effectiveness of interventions on parent-child relationships remains limited. Conversely, all interventions on teacher-child relationships were beneficial, with most results showing statistically significant outcomes for this subcategory. Only one measure was rated with moderate certainty, indicating a moderately sized beneficial effect from the EMR intervention; all other outcomes for teacher-child relationships were rated with very low certainty.

Transactional quality pertains to the indirect assessments by parents, peers, and teachers on factors affecting the quality of their interactions with target children. This domain had the fewest outcomes, further limiting the available information on intervention effectiveness in a key area of the transactional model of development. Similar to the relationship quality domain, intervention results concerning parents consisted solely of outcomes with very low certainty ratings. The intervention's impact on parent factors was entirely beneficial, with two outcomes showing statistically significant results.

Only one outcome contributed to understanding the effectiveness of interventions on peer factors, which greatly limited conclusions in this area. The SEAT intervention had a high certainty rating and showed a small adverse effect on peer factors. No other information about peer factors can be determined. Lastly, the effectiveness of interventions on teacher factors had four out of five outcomes rated with moderate to high certainty. Most interventions produced effects ranging from trivial to large beneficial impacts on teacher factors, three of which were statistically significant. Notably, EMR showed large beneficial effects despite low statistical power. KEY yielded inconsistent adverse and beneficial results, neither of which reached statistical significance.

Figure 6

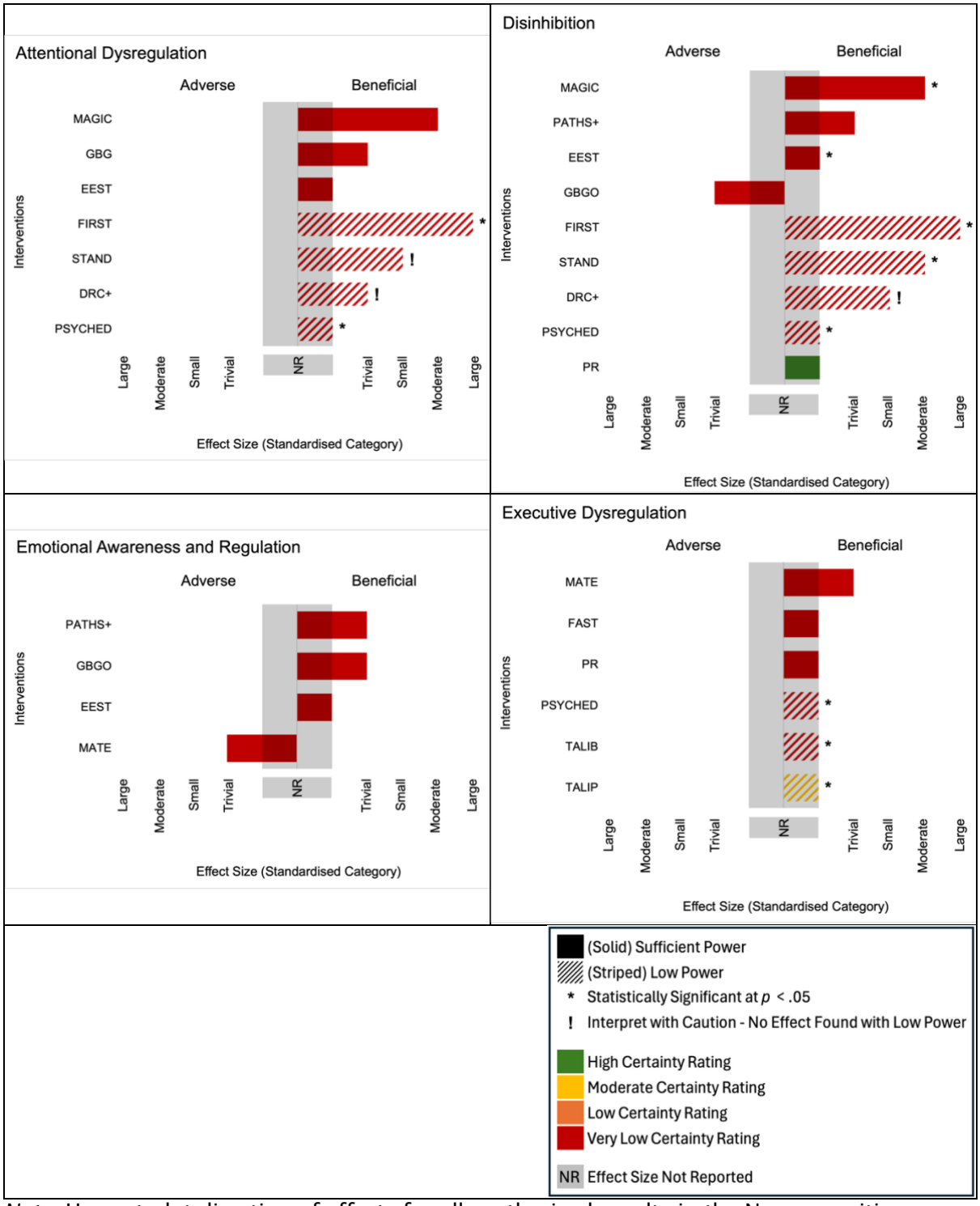
Direction of Effects for Academic Functioning



Note. Harvest plot direction of effects for all synthesised results in the Academic Functioning construct domain, split by construct subcategories.

Figure 7

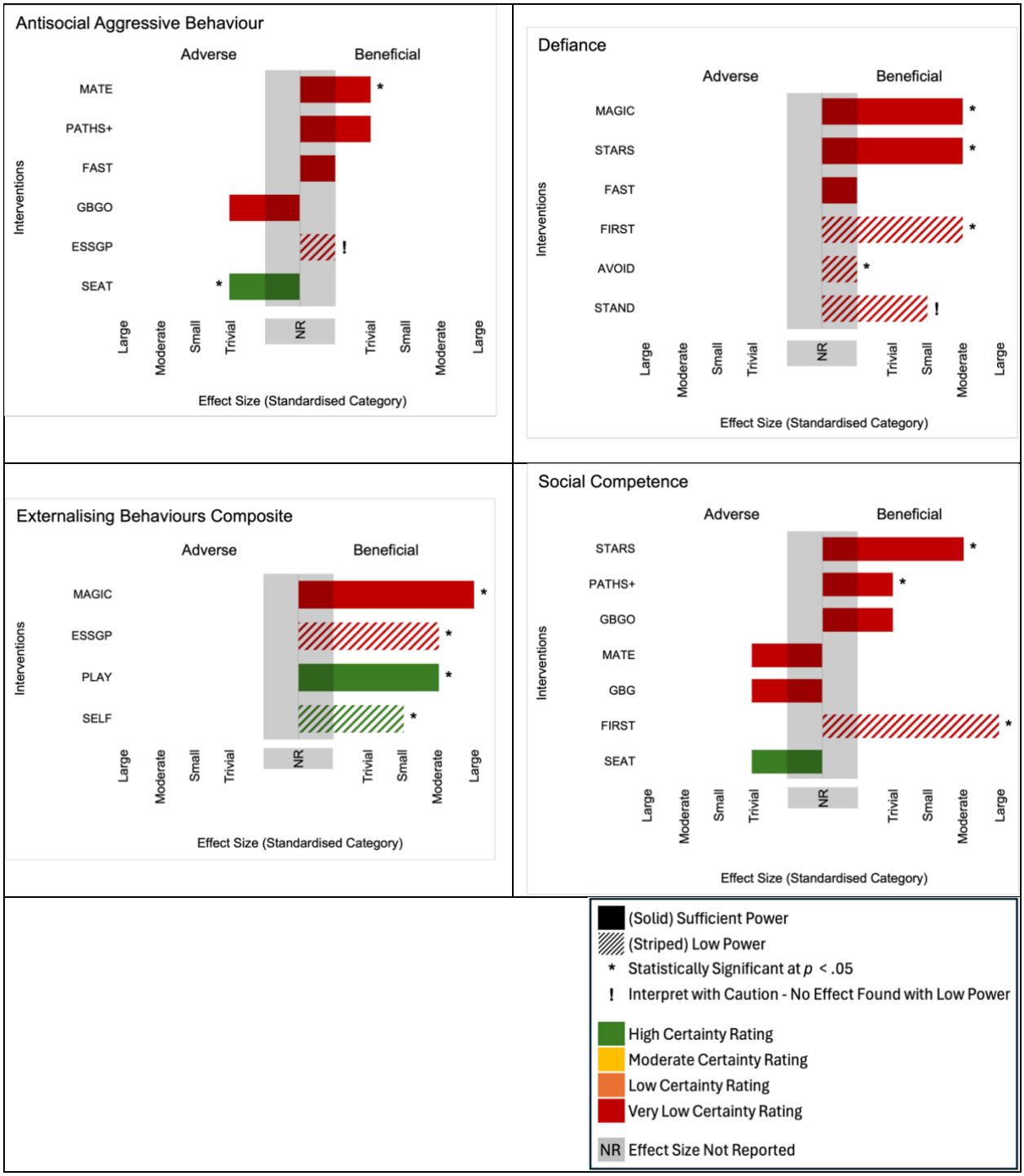
Direction of Effects for Neurocognitive Functioning



Note. Harvest plot direction of effects for all synthesised results in the Neurocognitive Functioning construct domain, split by construct subcategories.

Figure 8

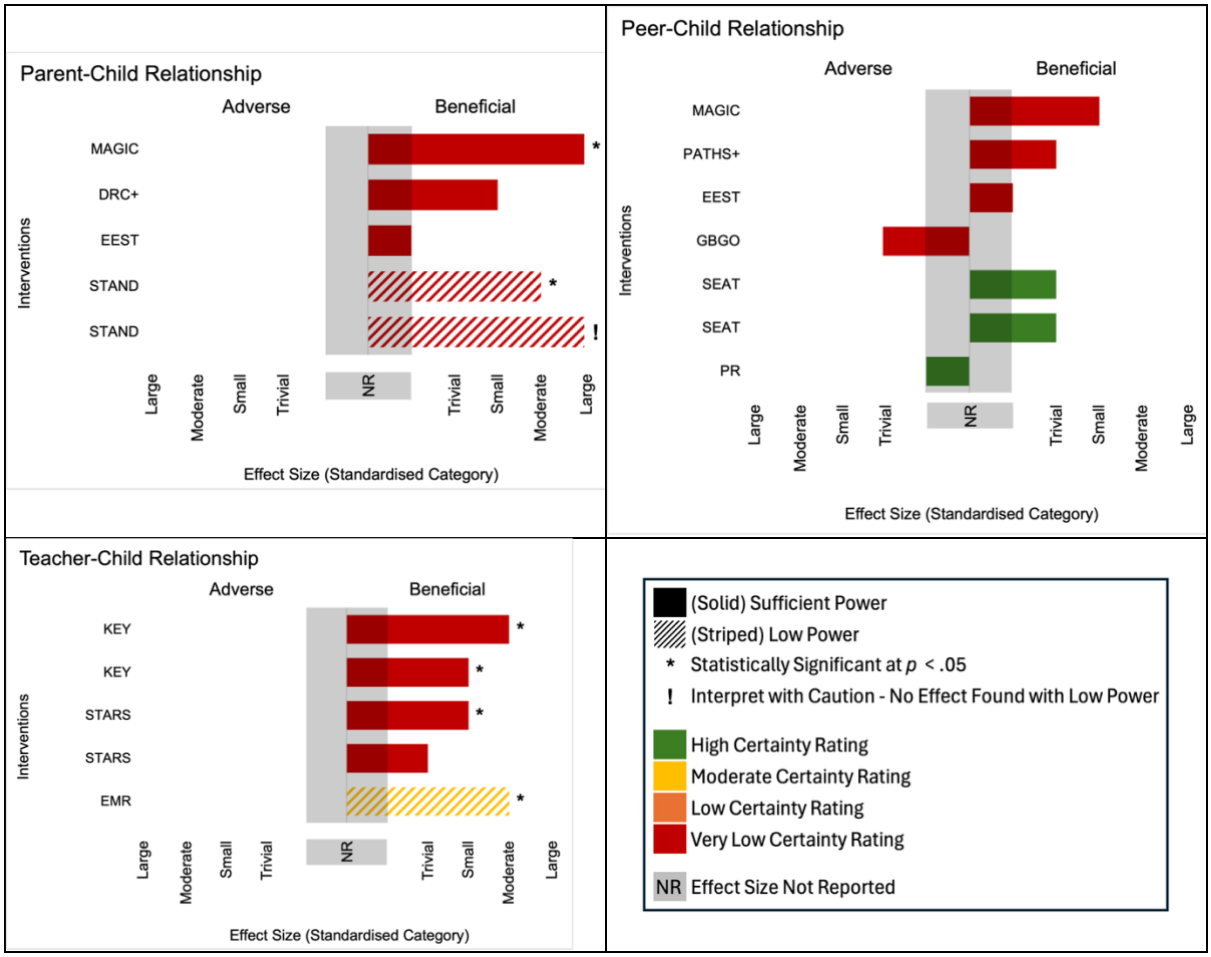
Direction of Effects for Social Functioning



Note. Harvest plot direction of effects for all synthesised results in the Social Functioning construct domain, split by construct subcategories.

Figure 9

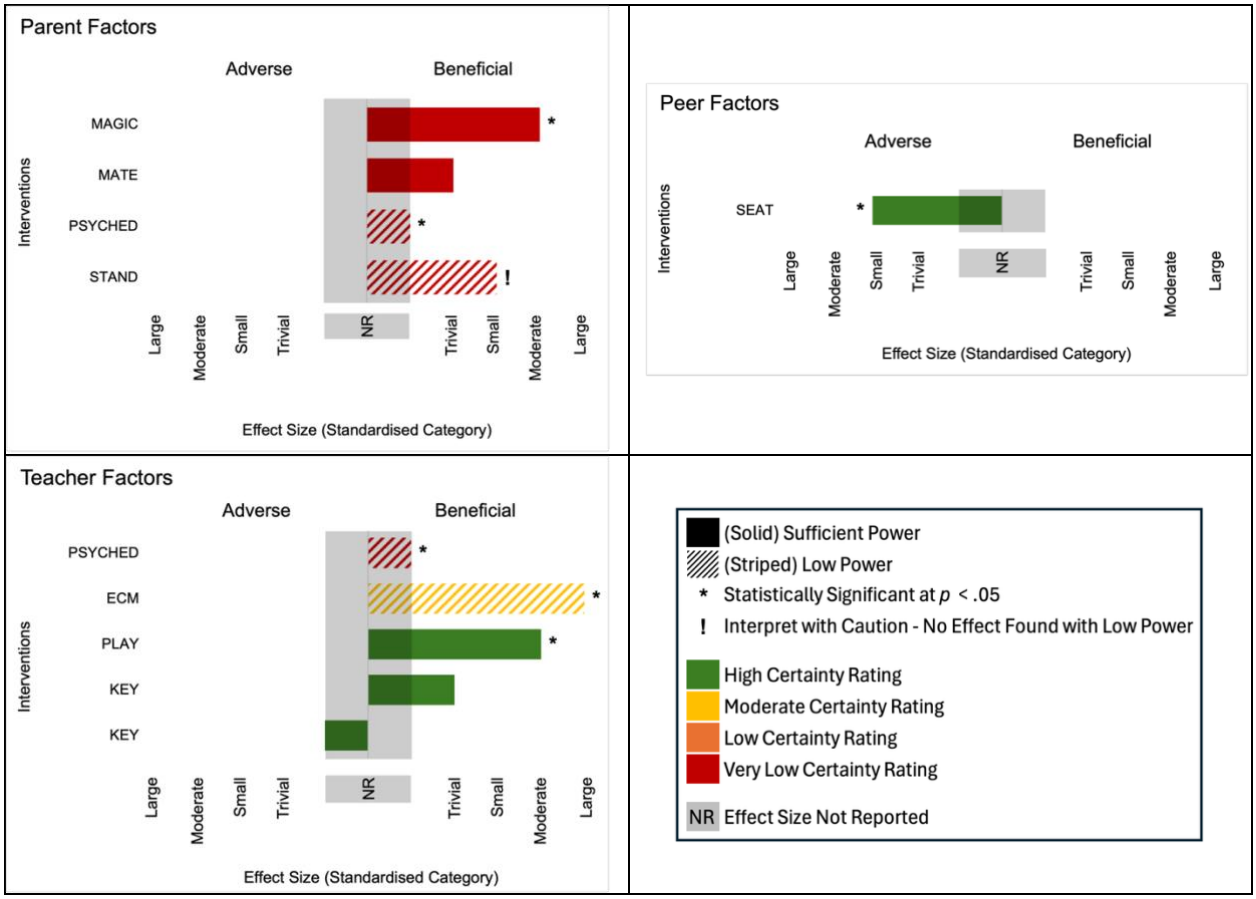
Direction of Effects for Relationship Quality



Note. Harvest plot direction of effects for all synthesised results in the Relationship Quality construct domain, split by construct subcategories.

Figure 10

Direction of Effects for Transactional Quality



Note. Harvest plot direction of effects for all synthesised results in the Transactional Quality construct domain, split by construct subcategories.

Sensitivity Analysis

To evaluate the effectiveness of interventions based on higher quality data, all outcomes with very low quality ratings were excluded in this sensitivity analysis (see Figure 11). Additionally, outcomes that did not show significant effects *and cam*e from adequately powered studies were also removed. Only nine of the 17 construct subcategories remained for the sensitivity analysis, and within these, the number of contributing study outcomes was notably reduced.

Academic functioning was reduced from three to two subcategories and from 20 contributing outcomes to nine. In academic functioning, all contributing outcomes from RATIO, GREET, EMR, STAND, and PLAY demonstrated beneficial effects, with nearly all outcomes showing a large effect.

Neurocognitive functioning was reduced from four to one subcategory and from 26 to one contributing outcome. In neurocognitive functioning, TALIP demonstrated a beneficial effect on executive dysregulation, although the reported effect size was not provided.

Social functioning was reduced from three to two subcategories, and from 23 to three contributing outcomes. In social functioning, PLAY and SELF showed small to moderate beneficial effects in composite measures of externalising behaviours. Alternatively, SEAT showed a trivial adverse effect on antisocial and aggressive behaviours.

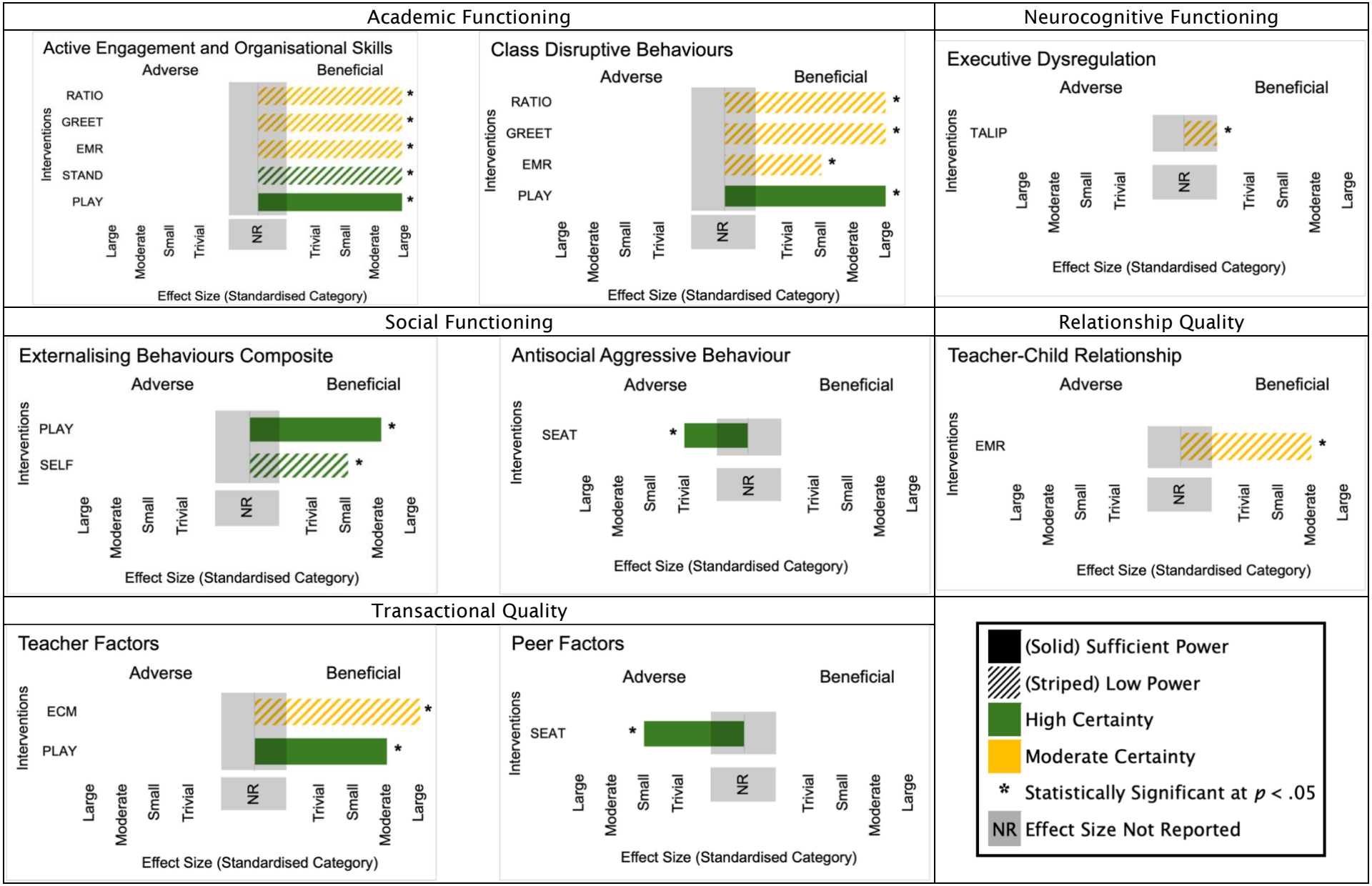
Relationship quality was reduced from three to one subcategory and from 17 to one contributing outcome. In terms of relationship quality, EMR demonstrated a moderate beneficial effect on teacher-child relationships.

Transactional quality was reduced from three to two subcategories and from 10 to three contributing outcomes. Notably, peer factors initially had only one outcome. In transactional quality, ECM and PLAY showed moderate to large beneficial effects on teacher factors. Conversely, SEAT exhibited a small adverse effect on peer factors.

Ultimately, the sensitivity analysis shows that there are not enough contributing outcomes from quality sources to confidently assess the effectiveness of interventions for externalising disorders.

Figure 11

Sensitivity Analysis of Synthesised Results



Moderators

Behavioural and Relational Domains

All outcome measures were reported and categorised into behavioural and relational domains, reflecting two main cyclical elements of transactional and developmental theories for externalising behaviours. Subcategories in the behavioural domain, 'Improve' and 'Reduce', were outcomes aimed at either enhancing or diminishing behavioural outcomes of the target children, respectively. Subcategories in the relational domain, 'Direct' and 'Indirect', represented outcomes that were either direct measures of the relationship between target children and related subgroups or indirect measures of factors influencing relationship quality.

Overarching Domains

Measures of behavioural and relational outcomes were disproportionate. Behavioural outcomes accounted for 67.57% of all outcomes. The weighting of behavioural to relational outcomes highlights a significant imbalance in the types of measures used to assess the effectiveness of interventions for externalising disorders in RCTs.

Behavioural and Relational Subcategories

While behavioural outcomes significantly surpass relational outcomes, the subcategories showed how researchers disproportionately used outcome measures *within* each domain. Negative behaviours, aimed to be reduced, represented the largest portion of outcomes targeted by studies in this review. Negative behaviours comprised 70% of behavioural outcomes and 47.30% of all outcome measures.

Positive behaviours, sought to be improved, comprised 30% of behavioural outcomes and 20.27% of all outcomes. Direct measures of relationships represented the smallest proportion of outcomes targeted by studies in this review. Direct measures of relationships accounted for 37.50% of relational outcomes and 12.16% of all outcomes. Indirect measures of relationships made up 62.50% of relational outcomes and 20.27% of all outcomes.

Overall, the included studies focused on measures of effectiveness on behaviours, specifically negative behaviours, and least on the quality of relationships in target children's lives. This weighting is not reflective of the dual nature-nurture influence involved in the development of externalising behaviours as outlined in developmental theories.

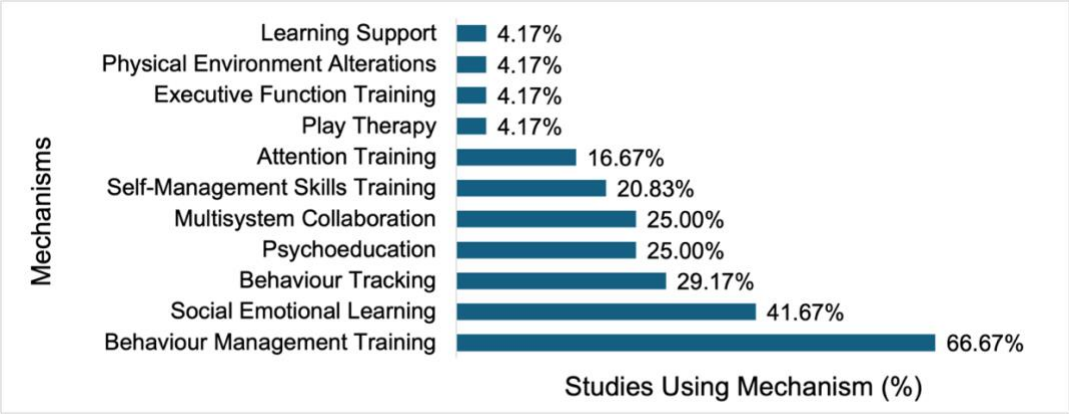
Mechanisms

Interventions involved between one and eight different combinations of ten mechanisms aimed at bringing about change. These mechanisms represent the underlying theoretical agents of change. Details on how each mechanism functions can be found in Appendix D. Due to the various combinations of mechanisms employed, there are too many confounding variables to identify clear patterns of intervention effectiveness linked to specific mechanisms. Instead, the patterns review in this moderator focused on the alignment between common mechanisms, risk and protective factors influencing the development of externalising disorders, and the outcome measures used. Aside from mechanisms related to medicinal, dietary, and sleep-based interventions—which were excluded from my review—the mechanisms included behaviour management training (BMT), social emotional learning (SEL), multisystem collaboration, behaviour tracking, psychoeducation, self-management skills training, attention training, play therapy, environmental strategies, and learning support.

With a strong focus on behavioural outcomes across the included studies, it is no surprise that the most common mechanism of change in interventions is behaviour management training, used by 66.67% of the included interventions (see Figure 12).

Figure 12

Percentage of Studies Using Each Mechanism



In light of the transactional model of development, targeting behaviours helps prevent escalation toward externalising disorders by promoting healthier transactions and the responses received from others. Almost half of the included studies involved elements

of SEL as mechanisms of change (41.67%). This mechanism supports both healthier behaviours and positive relational interactions through building awareness and teaching positive social and emotional skills. Outcome measures of behaviours to improve accounted for 20.27% of outcomes, while direct measures of relationships accounted for 12.16%. Compared to using SEL as a mechanism, measures of relationship quality as an indicator of intervention effects were disproportionately represented.

A matrix shows the studies that use specific mechanisms (see Figure 13). In this matrix, mechanisms are arranged from top to bottom by percentage of use, while interventions are ordered from left to right in decreasing volume of mechanisms used. At the lower end of use are play therapy, physical environmental changes, and learning support. Overall, teaching and learning form the foundation for most mechanisms, aiming to bring about change by developing knowledge and/or skills. Alternatively, multisystem collaboration, behaviour tracking, and physical environment changes offer additional or different ways to influence change. Multisystem collaboration explicitly seeks to effect change by bringing together people from multiple contexts to work towards co-created, targeted goals. Behaviour tracking impacts outcomes simply by increasing awareness and enabling repeated, real-time review of behaviour. Lastly, physical environment changes aim to bring about change through deliberate modifications to the physical setting.

Targeted-Subgroups

Interventions involved either single, dual, or multiple targets. A single target was the most common approach, making up 52% of interventions. Multisystem targets, which included more than two target subgroups, accounted for 32%. Dual targets were the least utilised, representing 16% of studies.

Figure 14 shows that interventions targeting a single subgroup had the highest proportion of significant beneficial effects (77.5%) with no significant adverse effects.

Figure 13

Mechanism Matrix

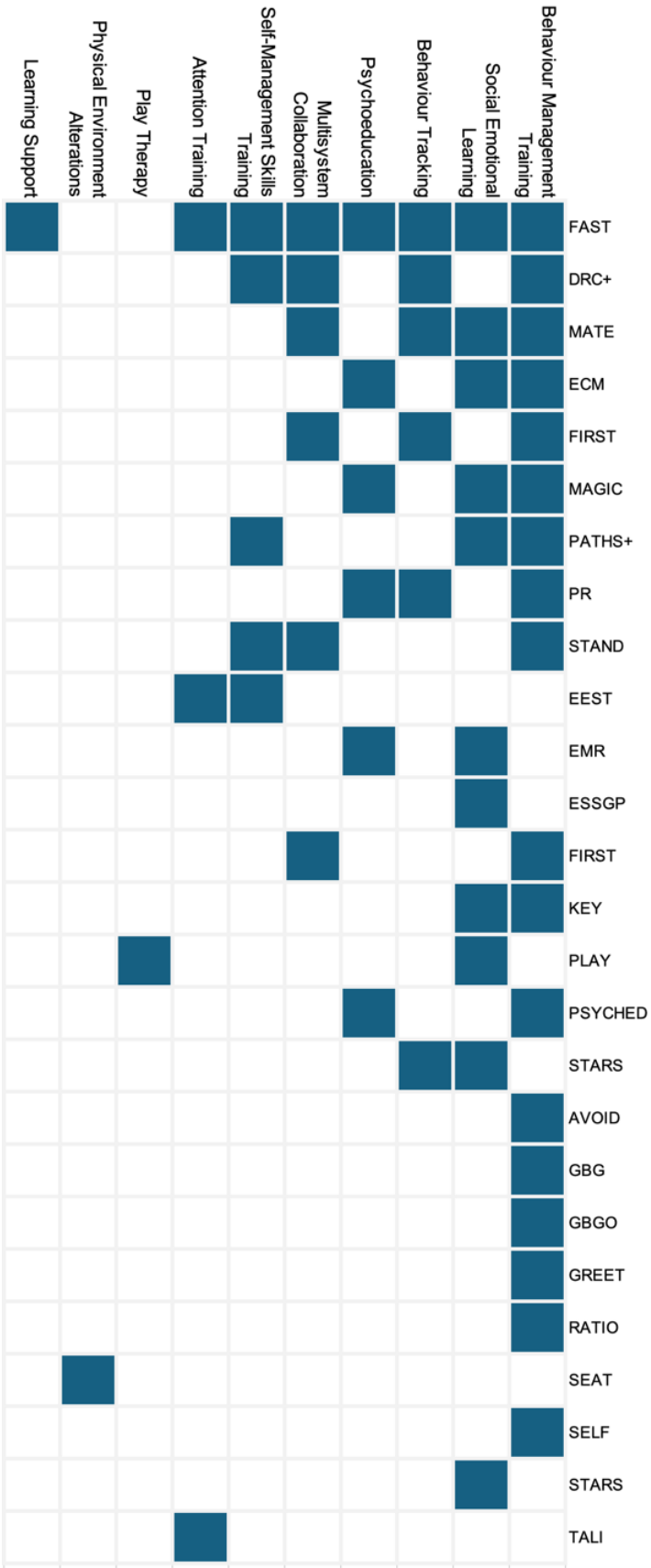


Figure 15 indicates that interventions targeting multisystem subgroups had the lowest proportion of significant beneficial effects (25%), with no significant adverse effects. Figure 16 demonstrates that interventions targeting dual-targeted subgroups produced significant beneficial effects in 45.83% of outcomes, and significant adverse effects in 8.33% of outcomes.

Target subgroups included children, peers, parents, and teachers. Children were the most targeted subgroup, accounting for 64% of interventions. Teachers were closely behind, targeted in 60% of interventions. Parents and peers were targeted half as often as children and teachers, with 36% and 32% of interventions respectively.

Figure 14

Synthesised Effects by a Single Targeted Subgroup

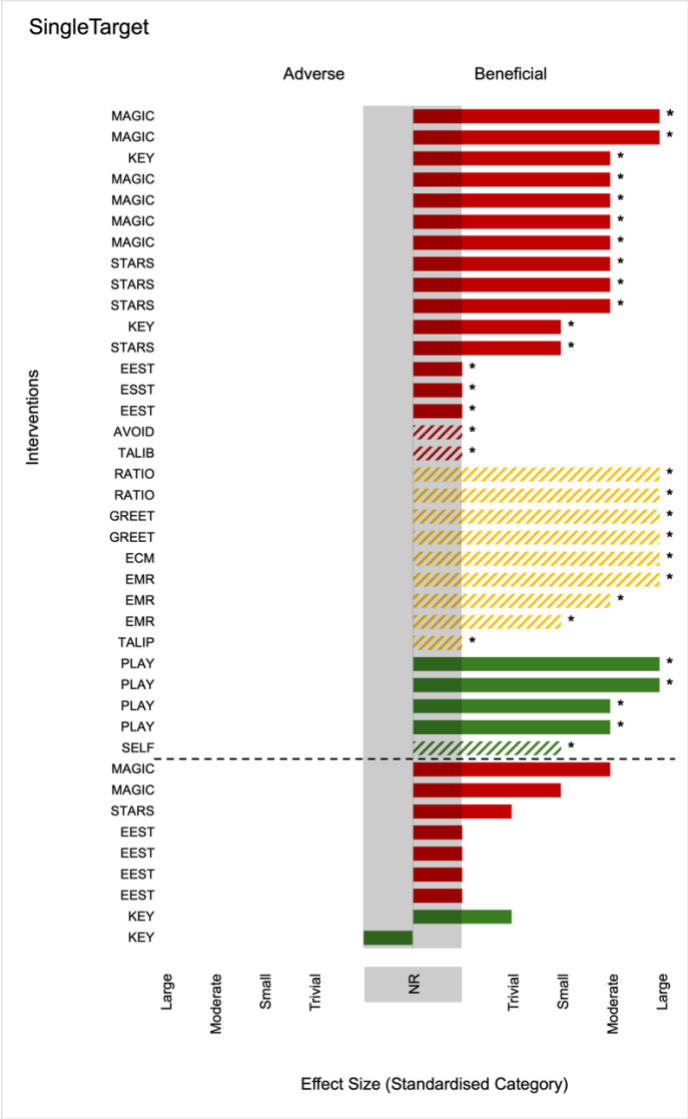


Figure 15

Synthesised Effects by Multisystem Targeted Subgroups

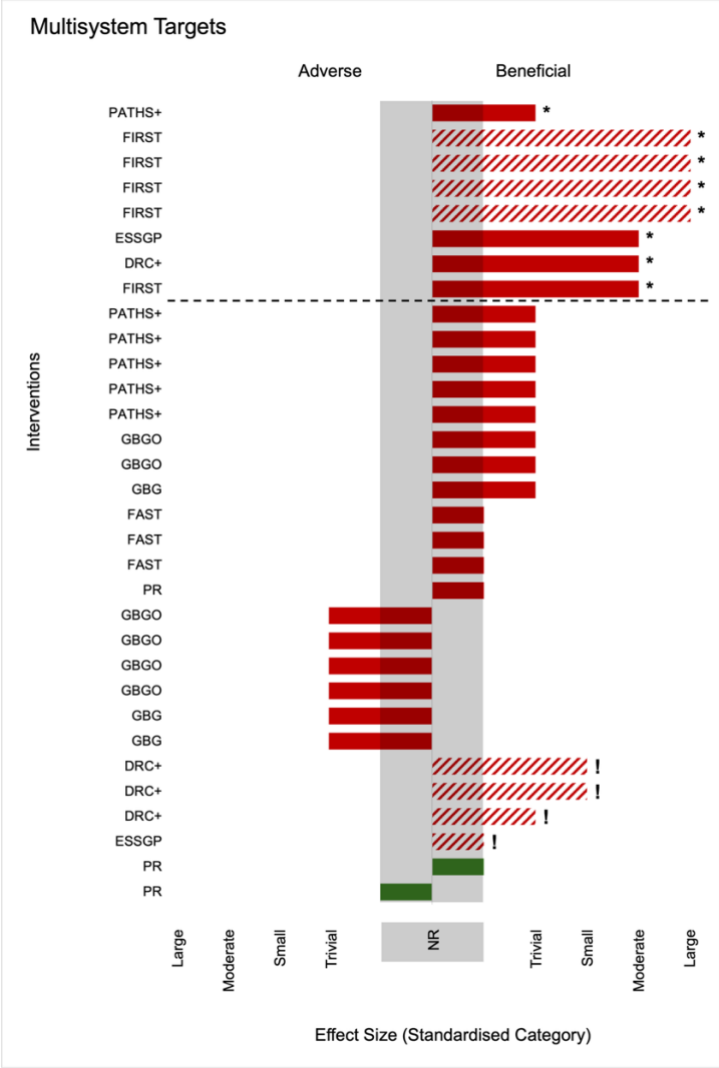
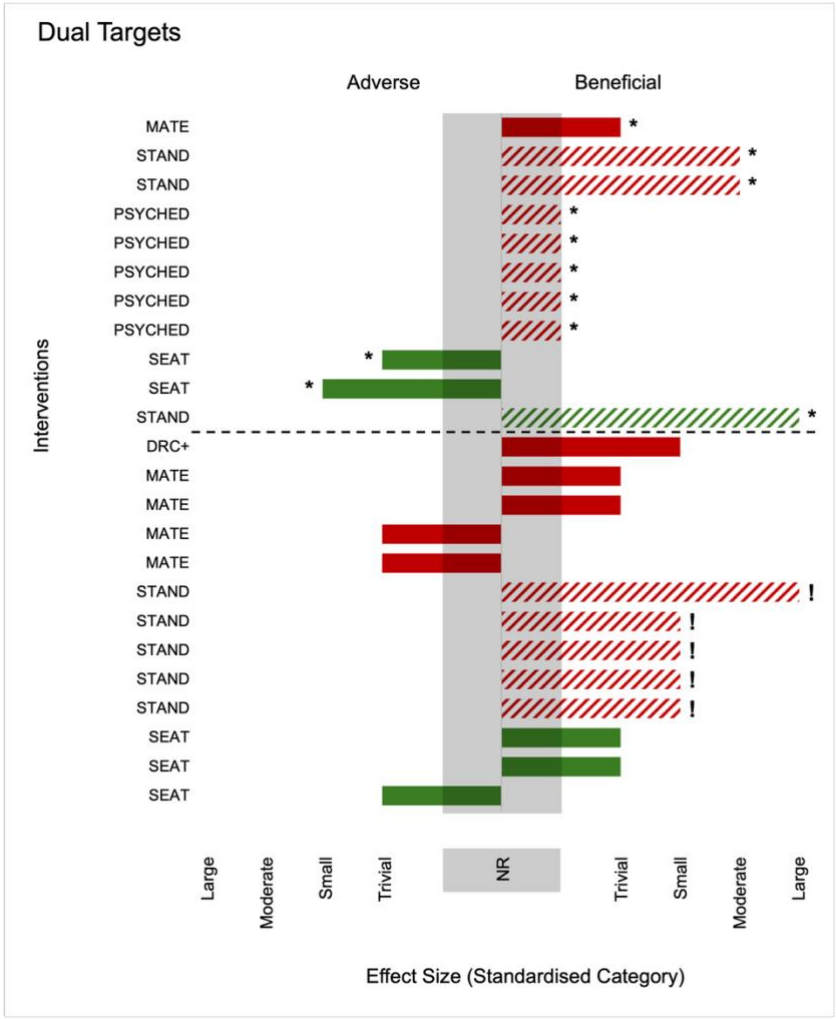


Figure 16

Synthesised Effects by a Dual-Targeted Subgroups



Certainty of Evidence

Construct subcategories had their certainty measured based on the proportions of contributing studies' certainty of evidence and the consistency of the effect direction within the construct subcategory. The effect direction and certainty of evidence are summarised in Table 5. Since all studies were RCTs, they all started with a rating of four. If a majority of contributing studies had high certainty ratings, the construct rating remained unchanged. If a majority of contributing studies were rated with moderate certainty and none were rated very low certainty then one point was deducted from the construct's certainty. If a majority of contributing studies were rated with moderate certainty and at least one was rated very low certainty, or if the number of high-rated studies equalled the combined total of moderate and very low-rated studies, then two points were deducted. Finally, if most contributing studies were rated with moderate certainty and very low certainty, and the number of very high ratings exceeded moderate ratings, three points were deducted. For consistency, if less than 70% of contributing studies aligned, one point was deducted. Peer factors were reduced post-hoc to low due to only having one contributing outcome within the construct subcategory.

As expected, almost all construct subcategories obtained a certainty rating of low or very low. So the direction of effect of reported results cannot be reviewed with much certainty at all. The quality of study methodology, in particular, the use of blinded raters, must be improved in research on the effectiveness of interventions for externalising disorders. The only construct subcategory with a high level of certainty was intervention effects on teacher factors that indirectly impact their relationship with target children. ECM and PLAY showed moderate to large beneficial impacts on teacher factors (see Figure 9).

Table 5

Summary of Evidence

SWiM Construct Subcategory	Direction of Effect	Study Type	Quality	Consistency	Certainty Rating	Certainty of Evidence
Behavioural Domain						
Academic Functioning						
Behaviours to Improve						
Active Engagement and Organisational Skills	Beneficial	4	-3	0	1	Very Low
Behaviours to Reduce						
Class Disruptive Behaviours	Beneficial	4	-3	0	1	Very Low
Engagement and Organisation Deficits	Beneficial	4	-3	0	1	Very Low
Neurocognitive Functioning						
Behaviours to Improve						
Emotional Awareness and Regulation	Beneficial	4	-3	0	1	Very Low
Behaviours to Reduce						
Attentional Dysregulation	Beneficial	4	-3	0	1	Very Low
Disinhibition	Beneficial	4	-3	0	1	Very Low
Executive Dysfunction	Beneficial	4	-3	0	1	Very Low
Social Functioning						
Behaviours to Improve						
Social Competence	Beneficial	4	-3	-1	0	Very Low
Behaviours to Reduce						
Antisocial Aggressive Behaviour	Beneficial	4	-3	0	1	Very Low
Defiance	Beneficial	4	-3	0	1	Very Low
Externalising Behaviours Composite	Beneficial	4	-2	0	2	Low
Relationship Quality						
Direct Measures of Relationships						
Parent-Child Relationship	Beneficial	4	-3	0	1	Low
Peer-Child Relationship	Adverse	4	-3	0	1	Low
Teacher-Child Relationship	Beneficial	4	-3	0	1	Low
Transactional Quality						
Indirect Measures of Relationships						
Parent Factors	Beneficial	4	-3	0	1	Low
Peer Factors	Adverse	4	0	0	4	Low
Teacher Factors	Beneficial	4	0	0	4	High

Discussion

Main Findings - The Point of RCTs

My review aimed to examine the effectiveness of interventions for primary-aged children exhibiting externalising behaviours across multiple settings. Regrettably, it was found that the quality of evidence in RCTs designed to objectively measure intervention effects in this area is compromised by poor study procedures, including the frequent use of unblinded raters—many of whom were participants in the studies. To put in the effort of conducting an RCT to gather objective evidence, only to then introduce rater bias by asking those very same people participating in the study for their opinion on their progress at post-test, is counterintuitive. Sibbald and Rolland (1998) note that RCTs are conducted because they have a greater capacity to rule out confounding variables than other types of research. They emphasise that researchers should "strive for methodological rigour and report their work in enough detail for others to assess its quality" (p. 201). The sensitivity analysis, which involved excluding studies of low quality, revealed that among all synthesised results, only a limited number of outcome measures remain suitable for reviewing intervention effects with an acceptable level of certainty. Only 17 of all 148 outcomes measured remained after excluding results with very low certainty, and just eight of these were rated with high certainty (see Figure 9). Furthermore, the presence of such low-quality evidence has hindered a thorough investigation of the moderating variables that influence intervention outcomes. I am hesitant to make any claims or interpretations of intervention effectiveness across the included studies, as the basis for any such claims is founded on poor-quality measures.

Most studies introduced systematic error through inadequate study procedures, thereby reducing the certainty of the reported results. The most common sources were unblinded outcome measurement, insufficient reporting of the randomisation processes, baseline imbalances, and high attrition. Specifically, 77.7% of the outcome measures included in my review were rated as having a high risk of bias due to the presence of unblinded raters. For instance, teacher-delivered interventions were often assessed using the unblinded participant teachers to complete outcome measures, introducing a risk of bias. To mitigate this, studies could have utilised independent observers or used objective behavioural coding systems. However, outcomes measured by objective devices or coding

systems were only used for 3 out of 148 outcome measures. These deficiencies lower confidence in reported effects. Many studies simply failed to report essential design elements required to assess study quality, such as allocation concealment, timing of outcome assessments, or handling of missing data. This lack of transparency meant I could not assess whether bias was introduced and therefore could not have high certainty in the reported results, which is particularly unfortunate for studies that had good procedures but omitted reporting them. This highlights the importance of clearly reporting study procedures for quality assessment and replicability.

The finding of methodological weaknesses, especially unblinded outcome assessment, has also been found in other reviews (Aldabbagh et al., 2024). Evidence suggests that inadequate blinding systematically distorts both beneficial and adverse effects. Trials with insufficient blinding have been found to inflate reports of beneficial outcomes (Hróbjartsson et al., 2013; Zhang et al., 2024).

A Necessary Pivot

Without the ability to speak to intervention effects with certainty, I returned to my objective of interpreting moderator effects on interventions and studies and reviewing how well they aligned with developmental theories related to externalising behaviours.

Accordingly, the discussion now shifts from reviewing "intervention effectiveness" to assessing how well the included interventions align with developmental theories of externalising disorders, particularly transactional and ecological models, using the moderators reported in the Results (mechanisms employed, targeted subgroups, and the balance of behavioural versus relational outcomes). In brief, alignment was strongest where interventions combined behavioural and relational components and engaged more than one target subgroup (child–teacher–parent), and weakest where studies targeted the child alone and measured only symptom reduction. The emphasis on behavioural outcome measures over relational measures and behaviour management over contextual change sits uneasily with Aotearoa longitudinal work on risk and protective factors for ODD and CD, which highlights the roles of caregiver–child relationships, peers, and the environment children grow up in (Fergusson et al., 2005). The remainder of the discussion therefore concentrates on moderators rather than aggregate effects: first, the distribution and implications of behavioural versus relational outcome measures; second, how the mechanisms interventions

most commonly use align with theory; and third, the types of subgroups targeted (single-agent, dual, multisystem) and what that implies for where studies are focused within an ecological systems framework. Throughout, the argument is that mechanisms and outcomes should be selected to represent various interchanging and cyclical influences that reflect the transactional model of development, rather than defaulting to the most easily observable indicator "of disruptive behaviour" – bad behaviour.

Behavioural and Relational Outcome Domains

A key finding of this moderator is the heavy over-representation of behavioural outcomes. One hundred of the 148 measured outcomes targeted behavioural domains, while relational outcomes were underrepresented. Although behavioural outcomes are easier to measure and code in RCTs, this bias may limit our understanding of holistic change. Across included studies, outcomes were disproportionately symptom-focused and focused on change in the individual child rather than relational or environmental change (e.g., parent–child relationship quality, school climate). Two studies illustrate the same pattern. Hoagwood et al. (2012) mapped child mental health outcomes and found that symptoms dominated measures of effectiveness, with an underrepresentation of environmental and context-based outcomes. Similarly, 94% of outcome measures in adolescent mental health studies measured symptoms. Other measures related to relationships or functioning were also underrepresented (Krause et al., 2019). Furthermore, measurement tool reviews show that outcome measures heavily rely on symptom American-based parent and teacher checklists to measure children's behaviour (Maldonado et al., 2019).

This imbalance likely reflects the feasibility and psychometric simplicity of scales as compared to the greater complexity of reliably capturing relationship quality (Deighton et al., 2014). A consequence of overly waiting for outcome measures towards behaviour and symptoms means we lose out on understanding various facets of intervention effectiveness that are important to the development of externalising disorders, and therefore limit research capacity to test theories of development. Shifting researchers of interventions in this area to require a more diverse set of core outcomes to measure will be beneficial. The ease of quantifying behavioural change is privileged over more nuanced relational outcomes, which are arguably central to interventions for EDs. Improved tools for valid,

objective relational measurement, such as coded video observations or teacher-student interaction ratings, could be further explored to reduce this gap.

Within the behavioural domain, outcomes were disproportionately categorised with 70% of behavioural outcomes being coded as Reduce (frequency/severity of problem behaviours) and 30% as Improve (prosocial/adaptive gains), favouring a symptom-decrease framework over a strength-building framework. There lies an irony in that developmental research strongly emphasises the role of positive reinforcement and encourages deliberately attending to and responding to positive behaviours (Moffat, 2011), yet outcome measures primarily focus on negative behaviours. Furthermore, not only did negative behaviours account for most behavioural outcomes, but they accounted for almost half of all outcome measures. So is there harm in measuring negative behaviours needing to be reduced in conditions related to negative behaviours? No, of course it is useful. However, it is harmful when studying the effects of interventions when outcomes are overly weighted toward one element of a condition. Particularly when the theorised causal factors are a balance between behavioural factors such as temperament, and relational factors such as positive transactions, attention can be biased toward positive or negative information so it must be a conscious choice to attend to the positive (Noguchi et al., 2006). Likewise, it research should balance its focus on positive outcomes. Are adaptive skills and emotional regulation skills not key skills to support healthy behaviours? The addition of positive skills is potentially even a more significant indicator of intervention effectiveness. Consider people who quit smoking – it is likely easier for the person who incorporates new, positive habits into their life, such as running, than for the person who stops with no skills or methods to help them stop. The value of measuring positive behaviours helps identify the growth and development of skills and knowledge interventions can have for children to build adaptive skills across both warm and harsh contexts.

Relational outcomes—such as warmth, conflict, or parent stress—were rarely measured. Only 12.16% of outcomes were direct relationship measures, and indirect relationship indicators (20.27%) were also underrepresented. Omitting relational and social domains in outcome measures is a missed opportunity, especially given a founding framework, Te Whare Tapa Whā, that highlights the benefits of addressing multiple domains, and evidence that interventions can improve peer relationships, empathy, and teacher–child

closeness (Durie, 2004). My review's findings further emphasise the need to recognise the interface between Western science and mātauranga Māori. Within te āo Māori, the centrality of whānau and whakawhanaungatanga reflects a deep epistemological understanding that relationships and community form the foundation of hauora and wellbeing. This holistic focus offers a clear advantage for understanding how to design and deliver effective interventions for children exhibiting externalising behaviours. By contrast, Western science has historically over-medicalised "bad behaviour," mainly placing responsibility and blame on the child or individual. Only in recent decades has developmental research begun to re-engage with transactional models of development and attachment theory, emphasising the interplay between relational environments and behavioural outcomes. The lack of focus on relational outcomes and community-building interventions across the reviewed RCTs sharply contrasts with kaupapa Māori approaches, such as Huakina Mai, which explicitly embed relationships and collective support at the centre of behaviour and wellbeing (Savage et al., 2014). Aligning Western evidence with Māori epistemologies at this interface offers significant potential for advancing interventions that address both behavioural challenges and the relational systems within which they develop.

When targeting behavioural outcomes, ADHD was frequently measured by standard assessments of hyperactivity and inattention, sometimes using instruments like cognitive attention and memory tests that may not always be generalizable to presentations of ADHD symptoms (Haywood et al., 2022), which can be affected by various brain networks such as the default mode network (Castellanos & Proal, 2012). Standard measures of ADHD are useful and facilitate easier comparison across many studies; however, alternative methods also exist for assessing ADHD in terms of its indirect impacts on quality of life and relationships (Mrug et al., 2001). Relational measures provide an additional valuable means of capturing the broader effects interventions might have on the indirect relational impacts of ADHD. The benefits of measuring relational outcomes add a more nuanced understanding of the impacts of interventions and target a major risk and protective factor in the development of externalising disorders involving functional or dysfunctional interactions.

Targeted Subgroups

Analysis of targeted subgroups highlighted clear patterns in the scope of intervention delivery. The majority of studies adopted a single-target design (52%), typically directed toward the child or the teacher, while dual-target approaches were least common (16%). Multisystem interventions were present in 32% of studies. Across all included studies, children (64%) and teachers (60%) were most frequently targeted, while parents (36%) and peers (32%) were incorporated at substantially lower rates.

Alignment with Social Cognitive Theory

Bandura's Social Cognitive Theory emphasises the reciprocal interaction of behaviour, cognition, and environment, with modelling and reinforcement as key mechanisms of learning. Given that peers and parents are central agents of modelling and reinforcement, their relatively low representation as intervention targets indicates partial alignment with this framework. Many child- and teacher-focused programmes sought to directly modify children's behaviours or teachers' classroom management, but fewer utilised peer modelling or parental reinforcement as intervention mechanisms. This is a notable gap, as deficits in social cognition and skill acquisition are robust predictors of externalising disorders (Baumel et al., 2021; Sharp et al., 2011; Tsomokos & Flouri, 2024). Underutilising parents and peers in interventions, influential sources of observational learning and reciprocal determinism are overlooked.

Alignment with the Transactional Model of Development

Sameroff's (2009) transactional model underscores how child characteristics and caregiver or teacher responses are mutually reinforced over time. This dynamic is especially relevant for externalising behaviours, where harsh or inconsistent responses can unintentionally reinforce oppositional patterns (Georgiou & Fanti, 2014; Serbin et al., 2015). While 60% of interventions targeted teachers and 36% targeted parents, these rates suggest that nearly half of interventions excluded caregivers entirely and two in five excluded teachers. Such omissions limit opportunities to disrupt negative reciprocal cycles. Moreover, the high reliance on child-only programmes reflects a medicalised framing of externalising disorders as child-centred problems, rather than recognising the bidirectional interplay between children and their social contexts. Interventions targeting dual subgroups are more consistent with the transactional model, but this was the smallest proportion of targeted subgroups.

Alignment with Ecological Systems Theory

Bronfenbrenner's (2000) ecological model places children as a single factor within a broader nested system including microsystems involving family, school, and peers, to broader macrosystems involving the broader community, socioeconomic climate, and cultural context children live within. Interventions in this review most often targeted individuals within the microsystem (i.e., the child or teacher) rather than addressing multiple interacting systems simultaneously. Restricting the majority of interventions to individuals and microsystems overlooks the cumulative risk and protective factors posed by overlapping systems such as poverty, peer deviance, or parent stress (Gluckman & Hayne, 2011; Poulton et al., 2015).

Synthesis

Taken together, the targeting patterns of reviewed interventions show partial but incomplete alignment with developmental theories. Social Cognitive Theory suggests that greater incorporation of parents and peers would strengthen opportunities for modelling and reinforcement; yet, these groups were the least frequently targeted. The transactional model indicates that long-lasting change requires addressing both the child and their immediate adult relationships. However, almost half of the included interventions did not involve parents, and child-centred framing of externalising behaviours was perpetuated through the high focus of interventions that solely targeted children. Ecological Systems Theory advocates for multi-system engagement across a child's home, school, and community contexts but the inclusion of macrosystem level contexts was non-existent. However, a quarter of interventions did target multi microsystems.

Overall, while some interventions aligned with transactional and ecological models, the dominant reliance on single-agent, child-focused targets reflects a gap between theoretical understanding of externalising disorders and the practical design of RCTs. Future trials would benefit from explicitly situating subgroup selection within developmental frameworks, ensuring that intervention targets mirror the known risk and protective factors identified in longitudinal studies of ODD and CD in Aotearoa and beyond.

Mechanisms

The mechanisms underpinning interventions for externalising disorders (EDs) reflect, to varying degrees, the broader risk and protective factors identified in developmental

psychopathology. Longitudinal studies in Aotearoa, such as the Christchurch Health and Development Study and the Dunedin Multidisciplinary Health and Development Study, have consistently demonstrated that early conduct problems are strong predictors of later adverse outcomes, including criminal offending, poor educational attainment, and mental health difficulties (Fergusson et al., 2005; Poulton et al., 2015). Fortunately, trajectories are not inevitable; protective factors mitigate risk and promote resilience. Identified protective factors, including healthy parental attachment, prosocial friendships, and academic engagement, can mitigate risk and promote resilience. Effective interventions should therefore seek to reduce risk exposures while simultaneously bolstering protective factors across developmental domains.

Included interventions mostly used behavioural management strategies, frequently focusing on reducing disruptive behaviour through reinforcement or classroom programmes. These methods are well-supported by evidence and highly effective for immediate behavioural improvements. However, their main limitation is their dominance in interventions. While behavioural outcomes were most commonly measured, behaviour management techniques were disproportionately emphasised, often overlooking other mechanisms that target upstream or contextual factors influencing EDs. Incorporating a broader range of mechanisms could improve long-term developmental outcomes.

A key gap was the lack of integration between neuropsychological and literacy-related factors. Research from Aotearoa indicates that reading comprehension serves not only as an academic skill but also as a developmental buffer with wide-reaching impacts. Rucklidge et al. (2009) demonstrated that, among incarcerated youth, very low reading comprehension significantly predicted recidivism, even after accounting for IQ and other risk factors. Improving comprehension by as little as one standard deviation reduced reoffending by roughly 30%, emphasising literacy as a protective factor against persistent antisocial behaviour. Given the strong connection between externalising behaviours and academic underachievement, adding literacy support to intervention programs for EDs could offer significant preventative advantages. Only one intervention, FAST, utilised learning support as a mechanism of change.

Similarly, protective mechanisms can be embedded in school environments. Yet environmental strategies were almost absent from the reviewed RCTs. Only one study

attempted to manipulate physical space, and in a counterintuitive direction by placing conflicting children closer together. This is striking given evidence that classroom climate and density directly affect behavioural outcomes. Considering how much time is spent in the classroom, overcrowded classrooms are ethically questionable environments for raising our future generations (Hanushek et al., 1999). Research has long shown that higher classroom density correlates with greater disruptive behaviour, while reductions in class size can improve engagement and prosocial conduct (Finn et al., 2003; Maxwell, 2003).

An analogy can be drawn with animal welfare. Stocking density in dairy farming is regulated to protect behavioural needs such as lying, grooming, and feeding; overstocking is associated with heightened aggression and stress (EFSA Panel on Animal Health and Animal Welfare (AHAW) et al., 2023; Fregonesi et al., 2007). Yet comparable safeguards for children's environments in schools are absent, despite evidence that high density increases conflict and undermines learning. Reducing students per resource (teachers, time, space) could be conceptualised as an environmental mechanism to prevent EDs in the same way stocking density protects welfare in animals. There is no legal limit for class sizes in Aotearoa. In contrast, early childhood education regulations in Aotearoa require a minimum amount of indoor and outdoor space per child, recognising the necessity of density limits for development. Extending similar principles to primary classrooms would align educational environments with the principles of developmental science.

Overall, the imbalance in mechanisms reveals a missed opportunity. Behavioural management is very effective and should stay central, but its emphasis suggests a gap in intervention strategies that incorporate other mechanisms. These would support evidence-based protective factors like literacy support, healthy learning environments, and social-emotional learning targeted at subgroups related to the children in question. When assessing intervention effectiveness, it is helpful to consider how well intervention strategies align with established developmental risk and protective factors, offering a more complete and theoretically grounded response to externalising disorders.

A Response to ERO

The findings of this review closely align with the issues raised in the ERO report (Education Review Office, 2024). Teachers in Aotearoa consistently report increasing disruptive behaviours, inadequate training, and limited access to specialist support. My

review adds to this picture by showing that although behavioural interventions are often effective, the evidence base is dominated by child-focused strategies, with limited attention to moderators such as relational outcomes, multi-system targeting, or environmental conditions. These gaps restrict our understanding of which approaches best address the complex drivers of challenging behaviour in schools.

A key contribution of this review is its alignment of moderators with developmental theories, offering a framework for addressing concerns raised by ERO. For example, relational outcomes were rarely measured, despite teachers identifying strained classroom relationships as a major burden, and only one intervention utilised an environmental-based mechanism, despite overcrowded classrooms being an issue impacting child behaviour in Aotearoa classrooms (Gerritsen, 2024). These findings highlight missed opportunities for aligning interventions with the realities faced by educators.

Most critically, my review demonstrates that the current RCT evidence encompassing a broad range of interventions generally has poor methodological rigour, with high risks of bias limiting confidence in the reported effects. Without rigorous study procedures, research provides incomplete and uncertain guidance. To address this gap, the Ministry of Education should consider investing in large-scale, high-quality RCTs that uphold methodological rigour while testing interventions grounded in developmental theory and ecological perspectives. Such investment would not only respond to local needs but also align with international best practice, where leading reviews have called for balanced core outcome measures and high-quality trials to advance the evidence base in this area (Deighton et al., 2014; Krause et al., 2019). Robust, well-designed trials are essential for providing the evidence urgently needed to inform responses to challenging behaviour and to support families, children, teachers and students in Aotearoa.

Strengths and Limitations

A key strength of this review is its strict adherence to PRISMA guidelines at every stage, which ensures transparency, reproducibility, and comprehensive coverage of the available literature. Another strength is the detailed review of moderators, interpreted through established developmental theories, enabling findings to be contextualised beyond just intervention effectiveness. This theoretical integration offers a deeper understanding of

how interventions may align with or overlook known risk and protective factors in the development of externalising disorders.

However, several limitations should be recognised. First, although another reviewer supported the initial screening process, data extraction and risk of bias assessments were conducted by a single reviewer. While this was mitigated by keeping detailed notes and clear documentation of decisions, involving two reviewers would have increased reliability. Second, the overall quality of the available evidence was low, with methodological weaknesses across most studies, which limited the certainty of findings and prevented firm conclusions about the effectiveness of the interventions.

Future Directions

Future research in this field should prioritise greater methodological rigour in the design and reporting of intervention studies. Large-scale, well-powered RCTs that follow established quality frameworks, such as RoB2, are needed to ensure that confounding variables are controlled and that results are reliable and interpretable. Improved transparency in reporting essential design elements, including randomisation procedures, blinding of outcome assessors, and handling of missing data, will be critical for strengthening the evidence base. Furthermore, future reviews would benefit from dual-reviewer extraction and bias assessment processes to reduce the risk of error or bias in interpretation. Strengthening study quality in these ways will be essential to generate high-certainty evidence that can meaningfully inform practice and policy for supporting children with externalising disorders.

References

Aldabbagh, R., Glazebrook, C., Sayal, K., & Daley, D. (2024). Systematic Review and Meta-Analysis of the Effectiveness of Teacher Delivered Interventions for Externalizing Behaviors. *Journal of Behavioral Education*, *33*(2), 233–274. https://doi.org/10.1007/s10864-022-09491-4

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders.* (5th, text revision edn). American Psychiatric Publishing. https://www.mredscircleoftrust.com/storage/app/media/DSM%205%20TR.pdf#page=1.00

Armstrong, M. E., Tornblad, M. K., & Jones, K. S. (2020). The accuracy of interrater reliability estimates found using a subset of the total data sample: A bootstrap analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1377–1382. https://doi.org/10.1177/1071181320641329

Ashworth, E., Humphrey, N., & Hennessey, A. (2020). Game over? No Main or subgroup effects of the good behavior game in a randomized trial in English primary schools. *Journal of Research on Educational Effectiveness*, *13*(2), 298–321.

Atherton, O. E., Lawson, K. M., Ferrer, E., & Robins, R. W. (2020). The role of effortful control in the development of ADHD, ODD, and CD symptoms. *Journal of Personality and Social Psychology*, *118*(6), 1226–1246. https://doi.org/10.1037/pspp0000243

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology*, *63*(3), 575–582. https://doi.org/10.1037/h0045925

Barrett, P., Davies, F., Zhang, Y., & Barrett, L. (2015). The impact of classroom design on pupils' learning: Final results of a holistic, multi-level analysis. *Building and Environment*, *89*, 118–133. https://doi.org/10.1016/j.buildenv.2015.02.013

Baumel, A., Mathur, N., Pawar, A., & Muench, F. (2021). Psychosocial interventions for children with externalized behavior problems: An updated meta-analysis of moderator effects. *Journal of Child and Family Studies*, *30*, 65–86.

Baumrind, D., & Black, A. E. (1967). Socialization practices associated with dimensions of competence in preschool boys and girls. *Child Development*, 291–327.

Bayliss, H. R., & Beyer, F. R. (2015). Information retrieval for ecological syntheses. *Research Synthesis Methods*, *6*(2), 136–148. https://doi.org/10.1002/jrsm.1120

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, *50*(2), 837–865.

Bezerra, C. T., Grande, A. J., Galvão, V. K., Santos, D. H. M. D., Atallah, Á. N., & Silva, V. (2022). Assessment of the strength of recommendation and quality of evidence: GRADE checklist. A descriptive study. *Sao Paulo Medical Journal*. https://doi.org/10.1590/1516-3180.2022.0043.r1.07042022

Bikic, A., Reichow, B., McCauley, S. A., Ibrahim, K., & Sukhodolsky, D. G. (2017). Meta-analysis of organizational skills interventions for children and adolescents with Attention-Deficit/Hyperactivity Disorder. *Clinical Psychology Review*, *52*, 108–123. https://doi.org/10.1016/j.cpr.2016.12.004

Blair, R. J., Zhang, R. u, Bashford-Largo, J., Bajaj, S., Mathur, A., Ringle, J., Schwartz, A., Elowsky, J., Dobbertin, M., Blair, K. S., & Tyler, P. M. (2021). Reduced neural responsiveness to looming stimuli is associated with increased aggression. *Social Cognitive and Affective Neuroscience*, *16*(10), 1091–1099. https://doi.org/10.1093/scan/nsab058

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2010a). Risk factors for conduct disorder and oppositional/defiant disorder: Evidence from a New Zealand birth cohort. *J Am Acad Child Adolesc Psychiatry*, *49*(11), 1125–1133. Medline (20970700). https://doi.org/10.1016/j.jaac.2010.08.005

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2010b). Risk factors for conduct disorder and oppositional/defiant disorder: Evidence from a New Zealand birth cohort. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(11), 1125–1133.

Boon, M. H., & Thomson, H. (2021). The effect direction plot revisited: Application of the 2019 Cochrane Handbook guidance on alternative synthesis methods. *Research Synthesis Methods*, *12*(1), 29–33. https://doi.org/10.1002/jrsm.1458

Bradley, M. C., & Mandell, D. (2005). Oppositional defiant disorder: A systematic review of evidence of intervention effectiveness. *Journal of Experimental Criminology*, *1*(3), 343–365. https://doi.org/10.1007/s11292-005-0062-3

Bradshaw, C. P., Schaeffer, C. M., Petras, H., & Ialongo, N. (2010). Predicting Negative Life Outcomes from Early Aggressive-Disruptive Behavior Trajectories: Gender Differences in Maladaptation Across Life Domains. *Journal of Youth and Adolescence*, *39*(8), 953–966. https://doi.org/10.1007/s10964-009-9442-8

Bradshaw, C. P., Shukla, K. D., Pas, E. T., Berg, J. K., & Ialongo, N. S. (2020). Using complier average causal effect estimation to examine student outcomes of the PAX good behavior game when integrated with the PATHS curriculum. *Administration and Policy in Mental Health and Mental Health Services Research*, *47*, 972–986.

Braga, T., Cunha, O., & Maia, Â. (2018). The enduring effect of maltreatment on antisocial behavior: A meta-analysis of longitudinal studies. *Aggression and Violent Behavior*, *40*, 91–100. https://doi.org/10.1016/j.avb.2018.04.003

Braun, S. S., van den Berg, Y. H. M., & Cillessen, A. H. N. (2020). Effects of a seating chart intervention for target and nontarget students. *Journal of Experimental Child Psychology*, *191*, 104742. https://doi.org/10.1016/j.jecp.2019.104742

Bronfenbrenner, U. (2000). Ecological Systems Theory. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology* (Vol. 3, pp. 129–133). Oxford University Press. https://doi.org/10.1037/10518-046

Buchanan, A., & Flouri, E. (2001). 'Recovery' after age 7 from 'externalising' behaviour problems: The role of risk and protective clusters. *Children and Youth Services Review*, *23*(12), 899–914. https://doi.org/10.1016/S0190-7409(01)00174-8

Burns, A. & Reporter. (2024, June 5). *Fewer NZers able to access specialist mental health support*. RNZ. https://www.rnz.co.nz/news/national/518664/fewer-nzers-able-to-access-specialist-mental-health-support

Burt, S. A., Krueger, R. F., McGue, M., & Iacono, W. (2003). Parent-Child Conflict and the Comorbidity Among Childhood Externalizing Disorders. *Archives of General Psychiatry*, *60*(5), 505–513. https://doi.org/10.1001/archpsyc.60.5.505

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H.

(2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *The BMJ*, *368*, l6890. https://doi.org/10.1136/bmj.l6890

Caspi, A., Henry, B., McGee, R. O., Moffitt, T. E., & Silva, P. A. (1995). Temperamental Origins of Child and Adolescent Behavior Problems: From Age Three to Age Fifteen. *Child Development*, *66*(1), 55–68. https://doi.org/10.2307/1131190

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of Genotype in the Cycle of Violence in Maltreated Children. *Science*, *297*(5582), 851–854. https://doi.org/10.1126/science.1072290

Castellanos, F. X., & Proal, E. (2012). Large-scale brain systems in ADHD: Beyond the prefrontal–striatal model. *Trends in Cognitive Sciences, Special Issue: Cognition in Neuropsychiatric Disorders*, *16*(1), 17–26. https://doi.org/10.1016/j.tics.2011.11.007

Chamberlain, M. C., & Medina, E. J. (2020). A case of being the same? Australia and New Zealand's reading in focus. *Australian Journal of Education*, *64*(3), 243–263. https://doi.org/10.1177/0004944120953235

Chesterfield, J. A., Porzig-Drummond, R., Stevenson, R. J., & Stevenson, C. S. (2021). Evaluating a Brief Behavioral Parenting Program for Parents of School-aged Children with ADHD. *Parenting*, *21*(3), 216–240. https://doi.org/10.1080/15295192.2020.1777783

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development*, *94*(5), 1181–1204. https://doi.org/10.1111/cdev.13968

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Conduct Problems Prevention Research Group. (2011). The Effects of the Fast Track Preventive Intervention on the Development of Conduct Disorder Across Childhood. *Child Development*, *82*(1), 331–345. https://doi.org/10.1111/j.1467-8624.2010.01558.x

Cook, C. R., Coco, S., Zhang, Y., Fiat, A. E., Duong, M. T., Renshaw, T. L., Long, A. C., & Frank, S. (2018). Cultivating Positive Teacher–Student Relationships: Preliminary Evaluation of the Establish–Maintain–Restore (EMR) Method. *School Psychology Review*, *47*(3), 226–243. https://doi.org/10.17105/SPR-2017-0025.V47-3

Cook, C. R., Fiat, A., Larson, M., Daikos, C., Slemrod, T., Holland, E. A., Thayer, A. J., & Renshaw, T. (2018). Positive Greetings at the Door: Evaluation of a Low-Cost, High-Yield Proactive Classroom Management Strategy. *Journal of Positive Behavior Interventions*, *20*(3), 149–159. https://doi.org/10.1177/1098300717753831

Cook, C. R., Grady, E. A., Long, A. C., Renshaw, T., Codding, R. S., Fiat, A., & Larson, M. (2017). Evaluating the Impact of Increasing General Education Teachers' Ratio of Positive-to-Negative Interactions on Students' Classroom Behavior. *Journal of Positive Behavior Interventions*, *19*(2), 67–77. https://doi.org/10.1177/1098300716679137

Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic Treatment: A Meta-Analysis of Outcome Studies. *Journal of Family Psychology*, *18*(3), 411–419. https://doi.org/10.1037/0893-3200.18.3.411

Dall, M., Fellinger, J., & Holzinger, D. (2022). The link between social communication and mental health from childhood to young adulthood: A systematic review. *Frontiers in Psychiatry*, *13*, 944815. https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2022.944815/pdf

de Jong, S. R., Van den Hoofdakker, B. J., Van Der Veen-Mulders, L., Veenman, B., Twisk, J. W., Oosterlaan, J., & Luman, M. (2023). The efficacy of a self-help parenting program for parents of children with externalizing behavior: A randomized controlled trial. *European Child & Adolescent Psychiatry*, *32*(10), 2031–2042.

Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: A review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health*, *8*(1), 14. https://doi.org/10.1186/1753-2000-8-14

Dishion, T. J., & Stormshak, E. A. (2007). *Intervening in children's lives: An ecological, family-centered approach to mental health care.* American Psychological Association.

Dodd, S., Clarke, M., Becker, L., Mavergames, C., Fish, R., & Williamson, P. R. (2018). A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *Journal of Clinical Epidemiology*, *96*, 84–92. https://doi.org/10.1016/j.jclinepi.2017.12.020

Döpfner, M., Breuer, D., Schürmann, S., Wolff Metternich, T., Rademacher, C., & Lehmkuhl, G. (2004). Effectiveness of an adaptive multimodal treatment in children with Attention-Deficit Hyperactivity Disorder-global outcome. *European Child & Adolescent Psychiatry*, *13*(Supplement), i117–i129. https://doi.org/10.1007/s00787-004-1011-9

Education Review Office. (2024). *Time to Focus: Behaviour in our Classrooms* (Evaluation Report Digital: 978-1-991126-02-3; pp. 1–158). Te Kāwanatanga o Aotearoa | New Zealand Government. https://evidence.ero.govt.nz/documents/time-to-focus-behaviour-in-our-classrooms/

EFSA Panel on Animal Health and Animal Welfare (AHAW), Nielsen, S. S., Alvarez, J., Bicout, D. J., Calistri, P., Canali, E., Drewe, J. A., Garin-Bastuji, B., Gonzales Rojas, J. L., Gortázar Schmidt, C., Herskin, M., Michel, V., Miranda Chueca, M. Á., Padalino, B., Roberts, H. C., Spoolder, H., Stahl, K., Velarde, A., Viltrop, A., … Winckler, C. (2023). Welfare of dairy cows. *EFSA Journal*, *21*(5), e07993. https://doi.org/10.2903/j.efsa.2023.7993

Eriksson, I., Cater, Å., Andershed, A.-K., & Andershed, H. (2011). What Protects Youths From Externalising and Internalising Problems? A Critical Review of Research Findings and Implications for Practice. *Australian Journal of Guidance and Counselling*, *21*(2), 113–125. https://doi.org/10.1375/ajgc.21.2.113

Every-Palmer, S., Grant, M. L., Thabrew, H., Hansby, O., Lawrence, M., Jenkins, M., & Romans, S. (2024). Not heading in the right direction: Five hundred psychiatrists' views on resourcing, demand, and workforce across New Zealand mental health services. *Australian & New Zealand Journal of Psychiatry*, *58*(1), 82–91. https://doi.org/10.1177/00048674231170572

Fergusson, D. M., John Horwood, L., & Ridder, E. M. (2005). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in

adulthood. *Journal of Child Psychology and Psychiatry*, *46*(8), 837–849. https://doi.org/10.1111/j.1469-7610.2004.00387.x

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "Why's" of Class Size: Student Behavior in Small Classes. *Review of Educational Research*, *73*(3), 321–368. https://doi.org/10.3102/00346543073003321

Fregonesi, J. A., Tucker, C. B., & Weary, D. M. (2007). Overstocking Reduces Lying Time in Dairy Cows. *Journal of Dairy Science*, *90*(7), 3349–3354. https://doi.org/10.3168/jds.2006-794

Georgiou, S. N., & Fanti, K. A. (2014). Transactional associations between mother-child conflict and child externalising and internalising problems. *Educational Psychology*, *34*(2), 133–153. https://doi.org/10.1080/01443410.2013.785055

Gerritsen, J. (2024, July 4). *New schools and classrooms urgently needed in high-growth areas, ministry warns*. RNZ. https://www.rnz.co.nz/news/national/521227/new-schools-and-classrooms-urgently-needed-in-high-growth-areas-ministry-warns?utm_source=chatgpt.com

Gluckman, P., & Hayne, H. (2011). Improving the transition: Reducing social and psychological morbidity during adolescence. *Auckland, NZ: Office of the Prime Minister's Science Advisory Committee*.

GRADE Working Group. (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal*, *328*(7454), 1490. https://doi.org/10.1136/bmj.328.7454.1490

Gray, M. R., & Steinberg, L. (1999). Unpacking Authoritative Parenting: Reassessing a Multidimensional Construct. *Journal of Marriage and Family*, *61*(3), 574–587. https://doi.org/10.2307/353561

Gresham, F. M., & Elliott, S. N. (1987). The relationship between adaptive behavior and social skills: Issues in definition and assessment. *The Journal of Special Education*, *21*(1), 167–181.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Hanushek, E. A., Mayer, S. E., & Peterson, P. (1999). The evidence on class size. *Earning and Learning: How Schools Matter*, 131–168.

Haywood, D., Baughman, F. D., Mullan, B. A., & Heslop, K. R. (2022). What Accounts for the Factors of Psychopathology? An Investigation of the Neurocognitive Correlates of Internalising, Externalising, and the p-Factor. *Brain Sciences*, *12*(4), 421. https://doi.org/10.3390/brainsci12040421

Health New Zealand, Te Whatu Ora. (2024, September 24). *Mental Health and Addiction Workforce Plan 2024 – 2027—Health New Zealand | Te Whatu Ora*. Health New Zealand | Te Whatu Ora. https://www.tewhatuora.govt.nz/publications/mental-health-and-addiction-workforce-plan-2024-2027

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents*. Guilford Press.

Heywood, C., & Fergusson, D. (2016). A pilot study of functional family therapy in New Zealand. *New Zealand Journal of Psychology*, *45*(3), 12–22.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds). (2022). *Cochrane Handbook for Systematic Reviews of Interventions (Version 6.3)*. Cochrane. https://training.cochrane.org/handbook

Hoagwood, K. E., Jensen, P. S., Acri, M. C., Olin, S. S., Lewandowski, R. E., & Herman, R. J. (2012). Outcome Domains in Child Mental Health Research Since 1996: Have They Changed and Why Does it Matter? *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(12), 1241-1260.e2. https://doi.org/10.1016/j.jaac.2012.09.004

Holmes, S. E., Slaughter, J. R., & Kashani, J. (2001). Risk Factors in Childhood That Lead to the Development of Conduct Disorder and Antisocial Personality Disorder. *Child Psychiatry and Human Development*, *31*(3), 183–193. https://doi.org/10.1023/A:1026425304480

Hoogendijk, K., Holland, J. G., Tick, N. T., Hofman, A. W., Severiens, S. E., Vuijk, P., Maras, A., & Van Veen, D. (2020). Effect of Key2Teach on Dutch teachers' relationships with students with externalizing problem behavior: A randomized controlled trial. *European Journal of Psychology of Education*, *35*, 111–135.

Hosseinnia, M., Mazaheri, M. A., & Heydari, Z. (2024). Educational intervention of parents and teachers for children with attention deficit hyperactivity disorder. *Journal of Education and Health Promotion*, *13*(1), 53.

Hróbjartsson, A., Thomsen, A. S. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., Ravaud, P., & Brorson, S. (2013). Observer bias in randomized clinical trials with measurement scale outcomes: A systematic review of trials with both blinded and nonblinded assessors. *CMAJ : Canadian Medical Association Journal*, *185*(4), E201–E211. https://doi.org/10.1503/cmaj.120744

Iznardo, M., Rogers, M. A., Volpe, R. J., Labelle, P. R., & Robaey, P. (2020). The Effectiveness of Daily Behavior Report Cards for Children With ADHD: A Meta-Analysis. *Journal of Attention Disorders*, *24*(12), 1623–1636. https://doi.org/10.1177/1087054717734646

Kapalka, G. M. (2006). Avoiding repetitions reduces ADHD children's management problems in the classroom. *Emotional and Behavioural Difficulties*, *10*(4), 269–279. https://doi.org/10.1177/1363275205058999

Kinnealey, M., Pfeiffer, B., Miller, J., Roan, C., Shoener, R., & Ellner, M. L. (2012). Effect of Classroom Modification on Attention and Engagement of Students With Autism or Dyspraxia. *The American Journal of Occupational Therapy*, *66*(5), 511–519. https://doi.org/10.5014/ajot.2012.004010

Kirk, H. E., Richmond, S., Cornish, K. M., & Spencer-Smith, M. (2022). The Role of Age, Cognitive Ability, and ADHD Symptoms on Outcomes of Attention Training in Primary School Children. *Journal of Cognitive Enhancement*, *6*(2), 170–182. https://doi.org/10.1007/s41465-021-00229-0

Kirk, H., Gray, K., Ellis, K., Taffe, J., & Cornish, K. (2017). Impact of Attention Training on Academic Achievement, Executive Functioning, and Behavior: A Randomized Controlled Trial. *American Journal on Intellectual and Developmental Disabilities*, *122*(2), 97–117. https://doi.org/10.1352/1944-7558-122.2.97

Klasen, M., Wolf, D., Eisner, P. D., Eggermann, T., Zerres, K., Zepf, F. D., Weber, R., & Mathiak, K. (2019). Serotonergic Contributions to Human Brain Aggression Networks. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00042

Klopfer, K. M., Scott, K., Jenkins, J., & Ducharme, J. (2019). Effect of Preservice Classroom Management Training on Attitudes and Skills for Teaching Children With Emotional and Behavioral Problems: A Randomized Control Trial. *Teacher Education and Special Education*, *42*(1), 49–66. https://doi.org/10.1177/0888406417735877

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Krause, K. R., Bear, H. A., Edbrooke-Childs, J., & Wolpert, M. (2019). Review: What Outcomes Count? Outcomes Measured for Adolescent Depression Between 2007 and 2017. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(1), 61–71. https://doi.org/10.1016/j.jaac.2018.07.893

Lamborn, S. D., Mounts, N. S., Steinberg, L., & Dornbusch, S. M. (1991). Patterns of Competence and Adjustment among Adolescents from Authoritative, Authoritarian, Indulgent, and Neglectful Families. *Child Development*, *62*(5), 1049–1065. https://doi.org/10.1111/j.1467-8624.1991.tb01588.x

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lewis, G. J., Asbury, K., & Plomin, R. (2017). Externalizing problems in childhood and adolescence predict subsequent educational achievement but for different genetic and environmental reasons. *Journal of Child Psychology and Psychiatry*, *58*(3), 292–304. https://acamh.onlinelibrary.wiley.com/doi/pdfdirect/10.1111/jcpp.12655?download=true

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., 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 Medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lukens, E. P., & McFarlane, W. R. (2004). Psychoeducation as Evidence-Based Practice: Considerations for Practice, Research, and Policy. *Brief Treatment and Crisis Intervention*, *4*(3), 205–225. https://doi.org/10.1093/brief-treatment/mhh019

Mabe, P. A., Turner, M. K., & Josephson, A. M. (2001). Parent Management Training. *Child and Adolescent Psychiatric Clinics of North America, Current Perspectives on Family Therapy*, *10*(3), 451–464. https://doi.org/10.1016/S1056-4993(18)30040-3

Maccoby, E. E. (1984). Children's Dispositions and Mother-Child Interaction at 12 and 18 Months: A Short-Term Longitudinal Study. *Developmental Psychology*, *20*(3), 459.

Makin, T. R., & Orban De Xivry, J.-J. (2019). Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *eLife*, *8*, e48175. https://doi.org/10.7554/eLife.48175

Maldonado, B. N., Chandna, J., & Gladstone, M. (2019). A systematic review of tools used to screen and assess for externalising behaviour symptoms in low and middle income settings. *Global Mental Health*, *6*, e13. https://doi.org/10.1017/gmh.2019.11

Manatū Hauora, Ministry of Health. (2024, November 19). *Annual Update of Key Results 2023/24: New Zealand Health Survey | Ministry of Health NZ*. Annual Data Explorer 2023/24: New Zealand Health Survey. https://www.health.govt.nz/publications/annual-update-of-key-results-202324-new-zealand-health-survey

Maxwell, L. E. (2003). Home and School Density Effects on Elementary School Children: The Role of Spatial Density. *Environment and Behavior*, *35*(4), 566–578. https://doi.org/10.1177/0013916503035004007

McIntosh, K., Chard, D. J., Boland, J. B., & Horner, R. H. (2006). Demonstration of Combined Efforts in School-Wide Academic and Behavioral Systems and Incidence of Reading and Behavior Challenges in Early Elementary Grades. *Journal of Positive Behavior Interventions*, *8*(3), 146–154. https://doi.org/10.1177/10983007060080030301

Meany-Walen, K. K. (2020). The Current State of Adlerian Play Therapy Published Research. *Journal of Individual Psychology*, *76*(2), 176–186. https://doi.org/10.1353/jip.2020.0005

Meany-Walen, K. K., Bratton, S. C., & Kottman, T. (2014). Effects of Adlerian Play Therapy on Reducing Students' Disruptive Behaviors. *Journal of Counseling & Development*, *92*(1), 47–56. https://doi.org/10.1002/j.1556-6676.2014.00129.x

Mingebach, T., Kamp-Becker, I., Christiansen, H., & Weber, L. (2018). Meta-meta-analysis on the effectiveness of parent-based interventions for the treatment of child externalizing behavior problems. *PloS One*, *13*(9), e0202855.

Moffat, T. K. (2011). *Increasing the Teacher Rate of Behaviour Specific*. *12*(1).

Moffitt, T. E., Houts, R., Asherson, P., Belsky, D. W., Corcoran, D. L., Hammerle, M., Harrington, H., Hogan, S., Meier, M. H., Polanczyk, G. V., Poulton, R., Ramrakha, S., Sugden, K., Williams, B., Rohde, L. A., & Caspi, A. (2015). Is Adult ADHD a Childhood-Onset Neurodevelopmental Disorder? Evidence From a Four-Decade Longitudinal Cohort Study. *American Journal of Psychiatry*, *172*(10), 967–977. https://doi.org/10.1176/appi.ajp.2015.14101266

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are Reading and Behavior Problems Risk Factors for Each Other? *Journal of Learning Disabilities*, *41*(5), 417–436. https://doi.org/10.1177/0022219408321123

Mrug, S., Hoza, B., & Gerdes, A. C. (2001). Children with Attention-Defecit/Hyperactivity Disorder: Peer Relationships and Peer-Oriented Interventions. *New Directions for Child and Adolescent Development*, *2001*(91), 51–78. https://doi.org/10.1002/cd.5

Mulder, R., Foulds, J., & Beaglehole, B. (2023). Time for action, not words: The urgent rebuilding of New Zealand's mental health workforce. *New Zealand Medical Journal*, *136*(1576), 8–10. https://doi.org/10.26635/6965.e1576

Murray, D. W., Rabiner, D., Schulte, A., & Newitt, K. (2008). *Feasibility and integrity of a parent–teacher consultation intervention for ADHD students*. *37*, 111–126.

Navarro-Soria, I., Servera, M., & Burns, G. L. (2020). Association of foster care and its duration with clinical symptoms and impairment: Foster care versus non-foster care comparisons with Spanish children. *Journal of Child and Family Studies*, *29*(2), 526–533.

Nightingale, A. (2009). A guide to systematic literature reviews. *Surgery (Oxford), Determining Surgical Efficacy*, *27*(9), 381–384. https://doi.org/10.1016/j.mpsur.2009.07.005

Noguchi, K., Gohm, C. L., & Dalsky, D. J. (2006). Cognitive tendencies of focusing on positive and negative information. *Journal of Research in Personality*, *40*(6), 891–910. https://doi.org/10.1016/j.jrp.2005.09.008

O'Connor, K. M., & Hayes, B. (2020). How effective are targeted interventions for externalizing behavior when delivered in primary schools? *International Journal of School & Educational Psychology*, *8*(3), 161–173. https://doi.org/10.1080/21683603.2018.1530157

Ogundele, M. O. (2018). Behavioural and emotional disorders in childhood: A brief overview for paediatricians. *World Journal of Clinical Pediatrics*, *7*(1), 9–26. https://doi.org/10.5409/wjcp.v7.i1.9

O'Hare, L., Biggart, A., Kerr, K., & Connolly, P. (2015a). A Randomized Controlled Trial Evaluation of an After-School Prosocial Behavior Program in an Area of Socioeconomic Disadvantage. *The Elementary School Journal*, *116*(1), 1–29. https://doi.org/10.1086/683102

O'Hare, L., Biggart, A., Kerr, K., & Connolly, P. (2015b). A randomized controlled trial evaluation of an after-school prosocial behavior program in an area of socioeconomic disadvantage. *The Elementary School Journal*, *116*(1), 1–29.

Olivier, B. (2004). Serotonin and Aggression. *Annals of the New York Academy of Sciences*, *1036*(1), 382–392. https://doi.org/10.1196/annals.1330.022

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Peng, P., & Miller, A. C. (2016). Does attention training work? A selective meta-analysis to explore the effects of attention training and moderators. *Learning and Individual Differences*, *45*, 77–87. https://doi.org/10.1016/j.lindif.2015.11.012

Pepler, D. J., King, G., Craig, W., Byrd, B., & Bream, L. (1995). The development and evaluation of a multisystem social skills group training program for aggressive

children. *Child and Youth Care Forum*, *24*(5), 297–313. https://doi.org/10.1007/BF02128600

Poulton, R., Moffitt, T. E., & Silva, P. A. (2015). The Dunedin Multidisciplinary Health and Development Study: Overview of the first 40 years, with an eye to the future. *Social Psychiatry and Psychiatric Epidemiology*, *50*(5), 679–693. https://doi.org/10.1007/s00127-015-1048-8

Powell, G. (2024). *Evidence-Based Interventions for Primary-aged Children with/exhibiting Externalising Disorders: A Systematic Literature Review*. https://osf.io/9afzx/

Qian, Y., Chen, M., Shuai, L., Cao, Q.-J., Yang, L., & Wang, Y.-F. (2017a). Effect of an ecological executive skill training program for school-aged children with attention deficit hyperactivity disorder: A randomized controlled clinical trial. *Chinese Medical Journal*, *130*(13), 1513–1520.

Qian, Y., Chen, M., Shuai, L., Cao, Q.-J., Yang, L., & Wang, Y.-F. (2017b). Effect of an Ecological Executive Skill Training Program for School-aged Children with Attention Deficit Hyperactivity Disorder: A Randomized Controlled Clinical Trial. *Chinese Medical Journal*, *130*(13), 1513–1520. https://doi.org/10.4103/0366-6999.208236

Raine, A. (2002). Biosocial studies of antisocial and violent behavior in children and adults: A review. *Journal of Abnormal Child Psychology*, *30*, 311–326.

Riise, E. N., Wergeland, G. J. H., Njardvik, U., & Öst, L.-G. (2021). Cognitive behavior therapy for externalizing disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology Review*, *83*, 101954. https://doi.org/10.1016/j.cpr.2020.101954

Ringwald, W. R., Forbes, M. K., & Wright, A. G. C. (2023). Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychological Medicine*, *53*(2), 533–546. https://doi.org/10.1017/S0033291721001902

Rivenbark, J. G., Odgers, C. L., Caspi, A., Harrington, H., Hogan, S., Houts, R. M., Poulton, R., & Moffitt, T. E. (2018). The high societal costs of childhood conduct problems: Evidence from administrative records up to age 38 in a longitudinal birth cohort. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(6), 703–710. https://doi.org/10.1111/jcpp.12850

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon & R. Lerner (Eds), *Handbook of Child Psychology: Social, Emotional, and Personality Development* (6th edn, Vol. 3, pp. 99–166). Wiley.

Sajadi, S., Raheb, G., Maarefvand, M., & Alhosseini, K. A. (2020). Family problems associated with conduct disorder perceived by patients, families and professionals. *Journal of Education and Health Promotion*, *9*, 184. https://doi.org/10.4103/jehp.jehp_110_20

Samek, D. R., & Hicks, B. M. (2014). Externalizing Disorders and Environmental Risk: Mechanisms of Gene-Environment Interplay and Strategies for Intervention. *Clinical Practice (London, England)*, *11*(5), 537–547. https://doi.org/10.2217/CPR.14.47

Sameroff, A. (2009). The transactional model. In *The transactional model of development: How children and contexts shape each other* (pp. 3–21). American Psychological Association. https://doi.org/10.1037/11877-001

Sameroff, A. (2010). A Unified Theory of Development: A Dialectic Integration of Nature and Nurture. *Child Development*, *81*(1), 6–22. https://doi.org/10.1111/j.1467-8624.2009.01378.x

Savage, C., Macfarlane, S., Macfarlane, A., Fickel, L., & Hēmi, H. T. (2014). Huakina Mai: A Kaupapa Māori Approach to Relationship and Behaviour Support. *The Australian Journal of Indigenous Education*, *43*(2), 165–174. https://doi.org/10.1017/jie.2014.23

Schünemann, H. J., Mustafa, R. A., Brozek, J., Steingart, K. R., Leeflang, M., Murad, M. H., Bossuyt, P., Glasziou, P., Jaeschke, R., Lange, S., Meerpohl, J., Langendam, M., Hultcrantz, M., Vist, G. E., Akl, E. A., Helfand, M., Santesso, N., Hooft, L., Scholten, R., … Guyatt, G. H. (2020). GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *Journal of Clinical Epidemiology*, *122*, 129–141. https://doi.org/10.1016/j.jclinepi.2019.12.020

Seeley, J. R., Small, J. W., Walker, H. M., Feil, E. G., Severson, H. H., Golly, A. M., & Forness, S. R. (2009). Efficacy of the First Step to Success intervention for students with attention-deficit/hyperactivity disorder. *School Mental Health*, *1*(1), 37–48.

Serbin, L. A., Kingdon, D., Ruttle, P. L., & Stack, D. M. (2015). The impact of children's internalizing and externalizing problems on parenting: Transactional processes and

reciprocal change over time. *Development and Psychopathology*, *27*(4pt1), 969–986. https://doi.org/10.1017/S0954579415000632

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of Mind and Emotion Regulation Difficulties in Adolescents With Borderline Traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(6), 563-573.e1. https://doi.org/10.1016/j.jaac.2011.01.017

Sibley, M. H., Pelham, W. E., Derefinko, K. J., Kuriyan, A. B., Sanchez, F., & Graziano, P. A. (2013). A pilot trial of Supporting Teens' Academic Needs Daily (STAND): A parent-adolescent collaborative intervention for ADHD. *Journal of Psychopathology and Behavioral Assessment*, *35*, 436–449.

Sterne, J. A. C., Higgins, J. P. T., Eldridge, S. M., Reeves, B. C., Savović, J., & Group, R. 2 D. (2019a). *Excel tool to implement RoB 2 for cluster-randomized trials*. https://www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-cluster-randomized-trials

Sterne, J. A. C., Higgins, J. P. T., Eldridge, S. M., Reeves, B. C., Savović, J., & Group, R. 2 D. (2019b). *Excel tool to implement RoB 2 (standard version)*. https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2

Sterne, J. A., Savovic, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., & others. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Taubner, S., White, L., O., Zimmermann, J., Fonagy, P., & Nolte, T. (2012). Mentalization Moderates and Mediates the Link between Psychopathy and Aggressive Behavior in Male Adolescents. *Journal of the American Psychoanalytic Association*, *60*(3), 605–612. https://doi.org/10.1177/0003065112450148

Theule, J., Wiener, J., Tannock, R., & Jenkins, J. M. (2013). Parenting Stress in Families of Children With ADHD: A Meta-Analysis. *Journal of Emotional and Behavioral Disorders*, *21*(1), 3–17. https://doi.org/10.1177/1063426610387433

Thompson, A. M. (2014a). A Randomized Trial of the Self-Management Training and Regulation Strategy for Disruptive Students. *Research on Social Work Practice*, *24*(4), 414–427. https://doi.org/10.1177/1049731513509691

Thompson, A. M. (2014b). A randomized trial of the self-management training and regulation strategy for disruptive students. *Research on Social Work Practice*, *24*(4), 414–427.

Thomson, H. J., & Thomas, S. (2013). The effect direction plot: Visual display of non-standardised effects across multiple outcome domains. *Research Synthesis Methods*, *4*(1), 95–101. https://doi.org/10.1002/jrsm.1060

Tsomokos, D. I., & Flouri, E. (2024). The role of social cognition in mental health trajectories from childhood to adolescence. *European Child & Adolescent Psychiatry*, *33*(3), 771–786.

Twisk, J., Bosman, L., Hoekstra, T., Rijnhart, J., Welten, M., & Heymans, M. (2018). Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications*, *10*, 80–85. https://doi.org/10.1016/j.conctc.2018.03.008

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How Many Studies Do You Need?: A Primer on Statistical Power for Meta-Analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. https://doi.org/10.3102/1076998609346961

Veenman, B., Luman, M., & Oosterlaan, J. (2017). Further insight into the effectiveness of a behavioral teacher program targeting ADHD symptoms using actigraphy, classroom observations and peer ratings. *Frontiers in Psychology*, *8*, 1157.

Venta, A., Sharp, C., Fonagy, P., & Fletcher, J. M. (2021). *Developmental psychopathology*. John Wiley & Sons.

Walter, U. M., & Petr, C. G. (2011). Best Practices in Wraparound: A Multidimensional View of the Evidence. *Social Work*, *56*(1), 73–80. https://doi.org/10.1093/sw/56.1.73

Wannarka, R., & Ruhl, K. (2008). Seating arrangements that promote positive academic and behavioural outcomes: A review of empirical research. *Support for Learning*, *23*(2), 89–93. https://doi.org/10.1111/j.1467-9604.2008.00375.x

Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Zhang, F., Zhu, Y., Zhao, S., Zhang, Q., Tao, H., Wu, Y., & Jia, P. (2024). Discordant Information on Blinding in Trial Registries and Published Research: A Systematic

Review. *JAMA Network Open*, *7*(12), e2452274.

https://doi.org/10.1001/jamanetworkopen.2024.52274

Appendix A

The PICO Framework Guide

POPULATION
TARGET AGE: ○ Children aged between 5-14 years **Include if:** • The mean(SD) age of *child* participants included are ≥5 and ≤14 years. **Exclude if:** • The mean(SD) age of *child* participants included are <5 or 14< years. NB: If addition/subtraction of the mean standard deviation brings the mean age outside of this age range it is excluded. I.e., If a study's mean age is 13 (SD = 1.4). The standard deviated mean range reaches 14.4, which is outside inclusion criteria. • Exclude if the study only includes a single gender.
TARGET GENDER/RACE/ETHNICITY: ○ All **Include if:** • The *child* participants include all genders/races/ethnicities. **Exclude if:** • The study involves an ex/inclusion criteria based on gender/race/ethnicity.
TARGET TRAITS: ○ Externalising disorders/behaviours ○ Disruptive disorders/behaviours ○ Antisocial traits/behaviours ○ ADHD ○ Conduct Disorder ○ Intermittent Explosive disorder ○ Oppositional defiant disorder ○ Children with ASD are included if disruptive behaviours were directly measured. **Include if:** • The eligibility criteria of samples requires children involved to have or exhibit at least one of these traits. **Exclude if:** • None of these traits are mentioned or required. • The sample inclusion criteria exclusively includes internalising disorders.
SUBGROUPS RELATED TO THE TARGET POPULATION: ○ Peers/Class ○ Schools ○ Parents/Guardians ○ Teachers/School counsellors ○ Multi-systems ○ Families **Include if:** • Studies targeting these subgroups also involve or are specifically connected to children who meet the aforementioned age, gender/race/ethnicity and traits criteria. • The outcome measures of these studies relate to target children's behaviour, or to the relationship between the subgroup and target children, or factors affecting this relationship. **Exclude if:** • The outcome measures are not related to target children's behaviour or the relationship/factors affecting the relationship between the subgroup and target children. • The inclusion criteria for the subgroup sample indirectly place restrictions on the inclusion of target children. E.g., only foster parents – limits to only target children with foster parents being included.

INTERVENTION

TARGET TYPE:

- Applicable and appropriate to deliver in school populations and/or settings.

Include if:

- Interventions focus on cognition, behaviour, social skills, relationships, emotional intelligence/regulation, or psychoeducation.

Exclude if:

- Interventions focus on nutrition, sleep or medicine.
- Telehealth based intervention.

TARGET DELIVERY:

- Possible to deliver without a health professional

Include if:

- Non-health professionals are able to train in and/or deliver the intervention.

Exclude if:

- Intervention delivery is dependent on a health professional.

COMPARISON

STUDY DESIGN:

- RCT with a sufficient control for the effects of natural progression and development over time.

Include if:

- The study is a Randomised Controlled Trial (RCT).
- The study compares an experimental to a control group.
- The control group receives treatment/care/business/services as usual etc (TAU/CAU/BAU/SAU), no treatment, attentional control (i.e., no treatment given but control group given the same amount of attention/time as the experimental group), waitlist, reading materials only.
- If it is a multi-armed study where the data of each experimental group is separately compared to the control group.
- ABAB reversal designs that involve 2 groups where a control group also receives the ABAB reversal.

Exclude if:

- Non-randomised, non-controlled study.
- The study is comparing two experimental groups with each other.
- The control group is receiving alternate treatment (other than attention and reading materials) specifically during the study.
- Multi-armed studies merge data of the different experimental groups when comparing to the control.
- Single group ABAB reversal design without a second control group receiving ABAB design.

OUTCOME COMPARISON:

- Group X Time. I.e., Group (experimental vs. control) X Time (pre vs. post).

Include if:

- The pre and post relevant outcome measures are compared across both the experimental and control groups.
- Cluster randomised trials use data analysis that accounts for comparison of nested data.

Exclude if:

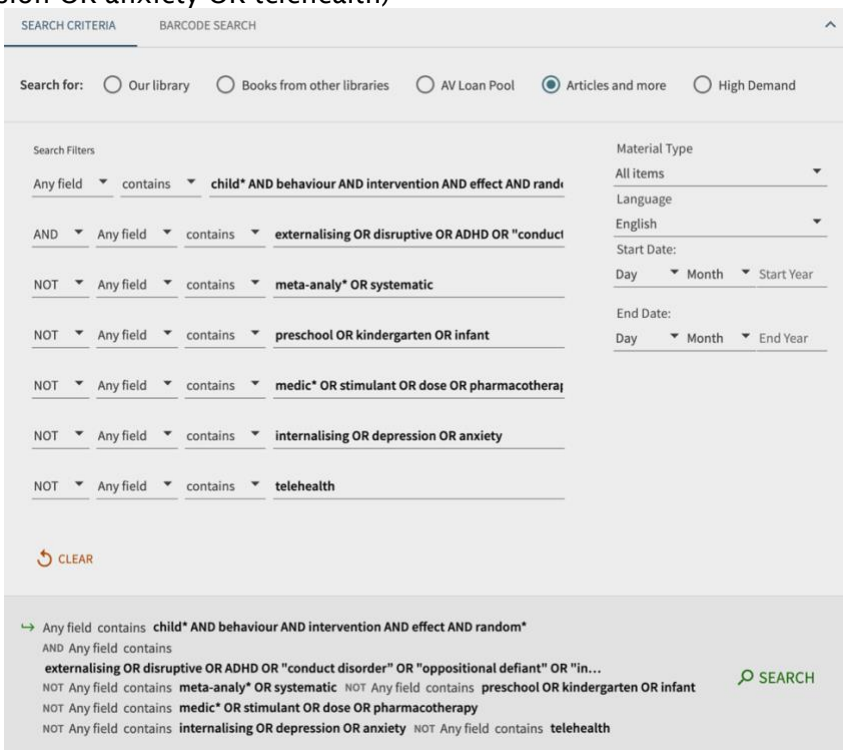
- The experimental and control groups pre and post measured outcomes are not compared.
- The comparison of nested data are not accounted for in the data analysis method.

OUTCOME
TARGET MEASURES: ○ BEHAVIOURAL: Increase/Decrease of anti/prosocial behaviours and traits related to Externalising/disruptive disorders. I.e., compliance, aggression etc ○ RELATIONAL: Changes in Relational warmth/conflict between the target population and related subgroups of the target population. **Include if:** • The outcome measures relate to the target traits aforementioned in the population PICO category (code as behavioural outcome). • The outcome measures relate to the relationship, or factors affecting the relationship, between the target population and the subgroups aforementioned. **Exclude if:** • Outcome measures are based on cognitive testing of attention, response time etc. • Outcome measures are perceptions *of theories* on behavioural/relational outcomes.

Appendix B

Line by Line Search Process

- Interface/Platform: The University of Waikato
- Date(s) accessed: 18/09/2024; 2/10/2024
- Who accessing: Gina Powell
- URL of source: https://waikato.primo.exlibrisgroup.com/discovery/search?vid=64WAIKATO_INST:64WAIKATO&lang=en&mode=advanced
- Line-by-Line search Process:
  1. Selected “Advanced Search”
  2. Under “Search for” I selected “Articles”
  3. Under “Material Type” I selected “Articles”
  4. Under “Language” I selected “English”
  5. Search Terms Used were: “(child* AND behaviour AND intervention AND effect AND random*) AND (externalising OR disruptive OR ADHD OR "conduct disorder” OR "oppositional defiant" OR "intermittent explosive" OR antisocial) NOT (meta-analy* OR systematic OR preschool OR kindergarten OR infant OR medic* OR stimulant OR dose OR pharmacotherapy OR internalising OR depression OR anxiety OR telehealth)”



  6. This search found 149 results before applying filters.
  7. This search was saved.
  8. Under “Availability” filter I included “Full text online” and “Peer-reviewed”.
  9. Under “Language” filter I included “English”and excluded “Japanese” and “Spanish”.
  10. This brought the search results to 119.
  11. All results were saved to favourites so they could be exported as one Excel file.
  12. The data was screened for duplicates in Excel by selecting the “Title” column, then selecting the Data tab and “Remove Duplicates” and “Expand the selection”, then deselecting “Select All” and reselecting “Title”. 1 duplicate value was found leaving **118** results.
  13. **I returned to the saved search in Discovery and substited the search term “intervention” for “treatment”.**
  14. This search found 74 results before applying filters.
  15. Under “Availability” filter I included “Full text online” and “Peer-reviewed”.
  16. Under “Language” filter I included “English”and excluded “Japanese”.
  17. This brought the search results to 60.
  18. I exported these results into excel, screened them for duplicates again. 1 duplicate value was found leaving **59** results.
  19. I pasted these 59 results to the other 118 and screened for duplicates again. 46 duplicate values were found leaving **131** results.

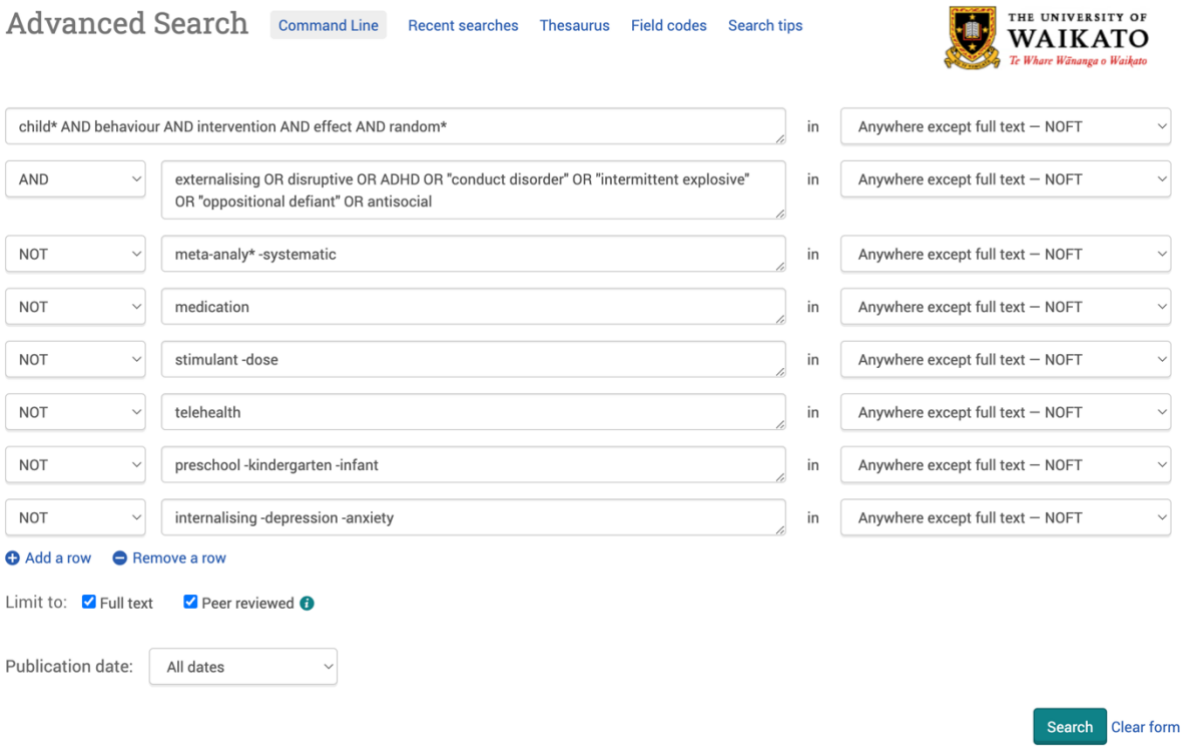
20. The “Title” and “description” columns of this excel spreadsheet were collated with the other databases’ Excel spreadsheets to begin the Screening Process document.

INNZ

- Interface/Platform: Index New Zealand (INNZ)
- Databases included:
- Limits applied: Full text and Peer reviewed.
- Date(s) accessed: 16/09/2024; 2/10/2024; 5/10/2024
- Who accessing: Gina Powell
- URL of source: https://natlib.govt.nz/collections/a-z/index-new-zealand-innz
- Line-byLine search Process:
    1. Selected “Advanced Search”
    2. Typed: “child* AND behaviour AND intervention” into the search bar
    3. Under “Availability” I selected “Only show items that are available online”.
    4. This search found 229 results before filters were applied.
    5. Under “Type” filter I selected “Journals”
    6. This search found 19 results.
    7. One of the results was excluded as it was a literature review. This brought the search results to **18**.
    8. All results were favourited and then the titles, descriptions, authors and year were manually entered into an Excel spreadsheet.
    9. **I returned to the original search in INNZ and substited the search term “intervention” for “treatment”.**
    10. This search found 89 results before applying filters.
    11. Under “Availability” filter I selected “Only show items that are available online”.
    12. Under “Type” filter I selected “Journals”.
    13. This brought the search results to **19**.
    14. I manually entered the titles, descriptions, authors and year of this new search into the same Excel spreadsheet leaving 37 results.
    15. The data was screened for duplicates in Excel by selecting the “Title” column, then selecting the Data tab and “Remove Duplicates” and “Expand the selection”, then deselecting “Select All” and reselecting “Title”. 4 duplicate values were found leaving **33** results.
    16. The “Title” and “Description” columns of this excel spreadsheet were collated with the other databases’ Excel spreadsheets to begin the Screening Process document.

Proquest

- Interface/Platform: Proquest
- Date(s) accessed: 1/10/2024; 2/10/2024; and 3/10/2024
- Who accessing: Gina Powell
- URL of source: https://www.proquest.com/
- Search Process:
    1. Selected “Advanced search”.
    2. Selected “Full text” and “Peer reviewed” boxes.
    3. Under “Source type” I selected “Scholarly Journals” and “Evidence-Based Medical Resources”
    4. Under “Document type” I selected “Article” and “Evidence Based Healthcare”.
    5. Under “Language” I selected “English”.
    6. Under “Result page options” I selected 100 items per page, and I selected “Exclude duplicate documents” box, include the “Spelling variants for your search terms” and “Form variants for your search terms”.
    7. Search terms used were: “noft(child* AND behaviour AND intervention AND effect AND random*) AND noft(externalising OR disruptive OR ADHD OR "conduct disorder" OR "intermittent explosive" OR "oppositional defiant" OR antisocial) NOT noft(meta-analy* -systematic) NOT noft(medication) NOT noft(stimulant -dose) NOT noft(telehealth) NOT noft(preschool -kindergarten -infant) NOT noft(internalising -depression -anxiety)”



8. This search found 393 results before applying filters.
9. Under 'Company/organisation' filter I excluded:
    i. American Academy of Sleep Medicine
    ii. Food & Drug Administration--FDA
    iii. International Bank for Reconstruction & Development--World Bank
    iv. National Institute on Drug Abuse
10. Under 'Database' filter I excluded:
    i. ABI/INFORM Collection
    ii. ABI/INFORM Global
    iii. ABI/INFORM Trade & Industry
    iv. Accounting, Tax & Banking Collection
    v. Advanced Technologies & Aerospace Database
    vi. Agriculture Science Database
    vii. Biological Science Database
    viii. Coronavirus Research Database
    ix. Engineering Database
    x. Environmental Science Database
    xi. Materials Science Database
    xii. Political Science Database
    xiii. Religion Database
    xiv. Research Library: Business
    xv. Research Library: History
    xvi. Research Library: Literature & Language
11. Databases *included* in the search were:
    i. Arts & Humanities Database
    ii. Australia & New Zealand Database
    iii. Canadian Business & Current Affairs Database: Health & Medicine
    iv. Canadian Business & Current Affairs Database
    v. Consumer Health Database
    vi. Criminal Justice Database
    vii. East & South Asia Database
    viii. East Europe, Central Europe Database
    ix. Education Database
    x. Health & Medical Collection
    xi. Healthcare Administration Database
    xii. Latin America & Iberia Database
    xiii. Library Science Database
    xiv. Middle East & Africa Database
    xv. Nursing & Allied Health Database
    xvi. ProQuest Central

xvii. ProQuest One Academic
xviii. Psychology Database
xix. Public Health Database
xx. Publicly Available Content Database
xxi. Research Library: Health & Medicine
xxii. Research Library: Social Sciences
xxiii. Research Library
xxiv. Science Database
xxv. Social Science Database
xxvi. Social Services Abstracts
xxvii. Sociological Abstracts
xxviii. Sociology Database
xxix. Turkey Database

12. Under "Publication title" filter I excluded:
    i. Archives of Disease in Childhood. Fetal and Neonatal Edition
    ii. BMC Palliative Care
    iii. Brain Impairment
    iv. Campbell Systematic Reviews
    v. International Journal of Mental Health and Addiction
    vi. International Journal of Preventive Medicine
    vii. Iranian Journal of Nursing and Midwifery Research
    viii. Journal of Aesthetics and Culture
    ix. Journal of Clinical Medicine
    x. Journal of the International AIDS Society
    xi. Microbial Ecology in Health and Disease
    xii. Molecular Psychiatry
    xiii. Psychiatry and Clinical Psychopharmacology
    xiv. Psychological Medicine
    xv. Psychopharmacology
13. Under "Language" filter I only included English
14. Under "Document type" I excluded "Literature Review" and included:
    i. Article
    ii. Feature
    iii. Evidence Based Healthcare
15. Under "Subject" filter I excluded:
    i. Anxiety
    ii. Autism
    iii. Drug Therapy
    iv. Drug Use
    v. Medical Diagnosis
    vi. Mental Depression
    vii. Meta-analysis
    viii. Pregnancy
    ix. Preschool Children
    x. Preschool Education
    xi. Stimulants
    xii. Systematic Review
16. This brought the search results to 111.
17. All results were exported as an Excel file.
18. The data was screened for duplicates in Excel by selecting the "Title" column, then selecting the Data tab and "Remove Duplicates" and "Expand the selection", then deselecting "Select All" and reselecting "Title". 2 duplicates were found leaving **109** results.

19. I returned to the saved search in Proquest and substited the search term "intervention" for "treatment".

20. This search found 370 results before applying filters.
    1. Under 'Company/organisation' filter I excluded:
        i. Allergan PLC
        ii. American Academy of Sleep Medicine
        iii. Food & Drug Administration--FDA
        iv. Institute of Medicine
        v. National Football League--NFL
        vi. Pfizer Inc
        vii. Sunovion Pharmaceuticals Inc
    2. Under 'Database' filter I excluded:
        viii. ABI/INFORM Collection

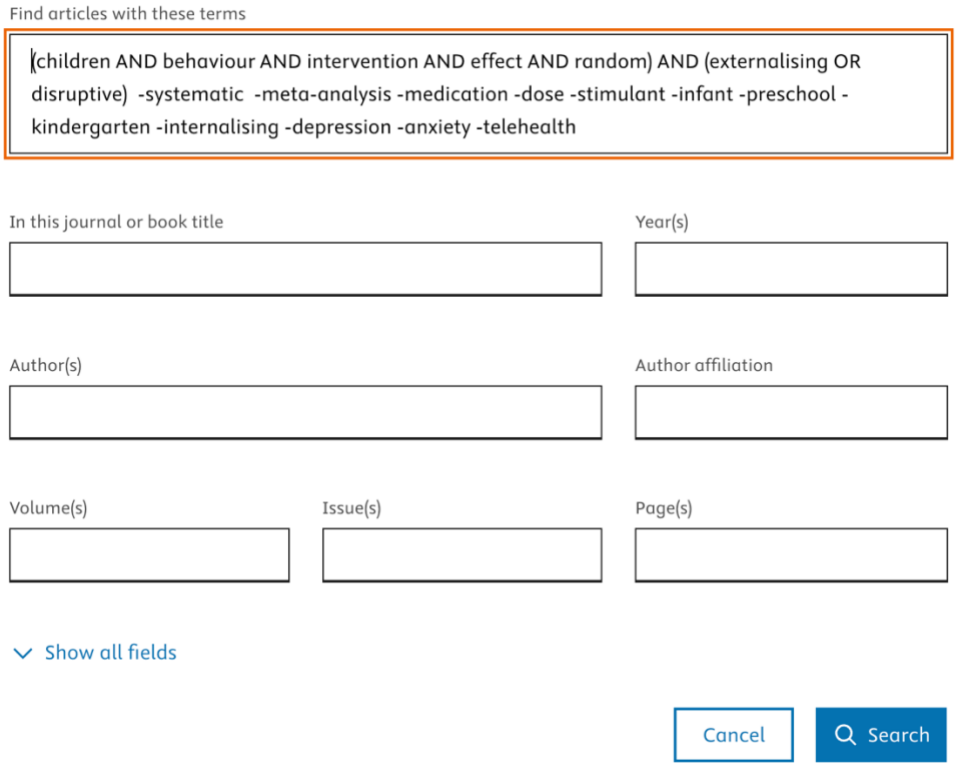
ix. ABI/INFORM Global
x. Advanced Technologies & Aerospace Database
xi. Agriculture Science Database
xii. Biological Science Database
xiii. Canadian Business & Current Affairs Database: Business
xiv. Coronavirus Research Database
xv. Earth, Atmospheric & Aquatic Science Database
xvi. Engineering Database
xvii. Environmental Science Database
xviii. Materials Science Database
xix. Political Science Database
xx. Research Library: Business
xxi. Research Library: History
xxii. Research Library: Literature & Language

3. Under "Publication title" filter I excluded:
   xxiii. Archives of Disease in Childhood. Fetal and Neonatal Edition
   xxiv. BMC Palliative Care
   xxv. Campbell Systematic Reviews
   xxvi. CNS Drugs
   xxvii. Journal of Child and Adolescent Psychopharmacology
   xxviii. Journal of Clinical Medicine
   xxix. Psychiatry and Clinical Psychopharmacology
   xxx. Psychological Medicine
   xxxi. Psychopharmacology
   xxxii. The International Journal of Neuropsychopharmacology
   xxxiii. Therapeutic Advances in Psychopharmacology
4. Under "Language" filter I only included English.
5. Under "Document type" I only included:
   xxxiv. Article
   xxxv. Feature
   xxxvi. Evidence Based Healthcare
6. Under "Subject" filter I excluded:
   xxxvii. Anxiety
   xxxviii. Autism
   xxxix. Central nervous system stimulants
   xl. Dopamine
   xli. Drug Therapy
   xlii. Drug Use
   xliii. Evidence-based medicine
   xliv. Medical Diagnosis
   xlv. Mental Depression
   xlvi. Meta-analysis
   xlvii. Methylphenidate
   xlviii. Pregnancy
   xlix. Preschool Children
   l. Preschool Education
   li. Psychotropic drugs
   lii. Stimulants
   liii. Systematic Review
7. This brought the search results to **74**.
8. I exported these results into excel, screened them for duplicates again. No duplicates were found.
9. I collated these 74 results to the other 109 and screened for duplicates again. 44 duplicate values were found leaving **139** results.
10. The "Title" and "description" columns of this excel spreadsheet were collated with the other databases' Excel spreadsheets to begin the Screening Process document.

Science Direct

- Interface/Platform: ScienceDirect
- Date(s) accessed: 16/09/2024; 3/10/2024
- Who accessing: Gina Powell

- URL of source: https://www.sciencedirect.com/
- Search Process:
    1. Selected “Advanced search”.
    2. Search terms used were: “(children AND behaviour AND intervention AND effect AND random) AND (externalising OR disruptive) -systematic -meta-analysis -medication -dose -stimulant -infant -preschool -kindergarten -internalising -depression -anxiety -telehealth”



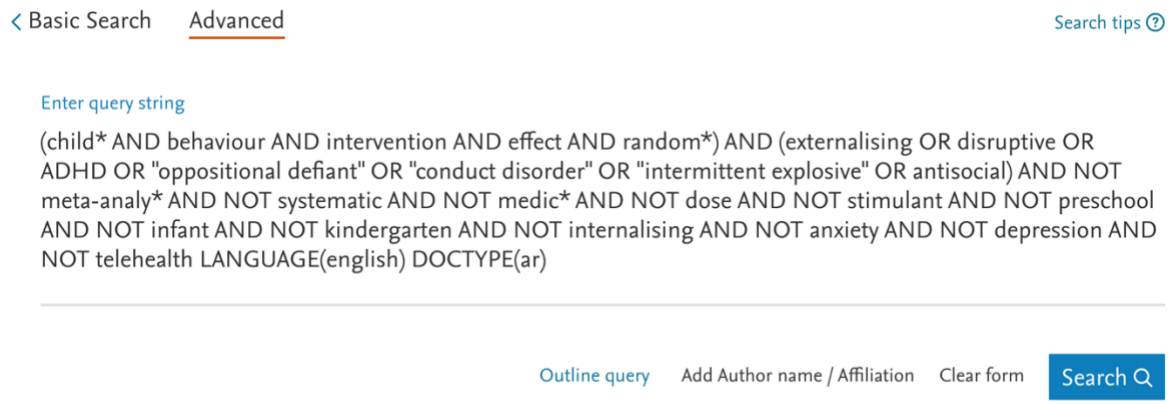
    3. This search found 2,431 results before applying filters.
    4. Under “Article Type” filter I selected “Research Articles”.
    5. Under “Publication Title” filter I included:
        i. Children and Youth Services Review
        ii. Social Science & Medicine
        iii. World Development
        iv. Research in Developmental Disabilities
        v. Child Abuse & Neglect
        vi. Journal of School Psychology
        vii. Journal of Criminal Justice
        viii. Computers & Education
        ix. Journal of Adolescence
        x. Computers in Human Behaviour
        xi. Health & Place
        xii. Social Science Research
        xiii. Personality and Individual Differences
        xiv. Teaching and Teacher Education
        xv. Heliyon
        xvi. Behavioural Brain Research
    6. Under “Subject areas” filter I selected:
        i. Social Sciences
        ii. Psychology
        iii. Medicine and Dentistry
        iv. Neuroscience
    7. This brought the search results to **331**.
    8. With 100 results per page, four pages were exported in ris format to Endnote. In Endnote I filed the results under a ScienceDirect Intervention folder.
    9. **I returned to the saved search in ScienceDirect and substited the search term “intervention” for “treatment”.**
    10. This search found 3,496 results before applying filters.
    11. Under “Article Type” filter I selected “Research Articles”.
    12. Under “Publication Title” filter I included:
        i. Children and Youth Services Review

ii. Social Science & Medicine
iii. Brain Research
iv. Child Abuse & Neglect
v. Science of the Total Environment
vi. Research in Developmental Disabilities
vii. World Development
viii. Neuropsychologia
ix. Physiology & Behavior
x. Journal of School Psychology
xi. Journal of Criminal Justice
xii. Behaviour Research and Therapy
xiii. Behavioural Brain Research
xiv. Computers in Human Behaviour
xv. Journal of Experimental Child Psychology

13. Under "Subject areas" filter I selected:
    i. Psychology
    ii. Social Sciences
    iii. Medicine and Dentistry
    iv. Neuroscience
14. This brought the search results to **292**.
15. With 100 results per page, three pages were exported in ris format to Endnote. In Endnote I filed the results under a ScienceDirect Treatment folder.
16. In Endnote I created a ScienceDirect Merged folder and added both sets of intervention (331) and treatment (292) results here leaving 623 results.
17. The merged folder was screened for duplicates in Endnote by selecting "Library" and "Find Duplicates" and "Keep Oldest". 134 duplicates were found and removed leaving **489** results.
18. From Endnote the results were exported as an xml file to be useable in Excel which created duplicates.
19. I exported these results into Excel, screened them for duplicates again. 3160 duplicate values were found leaving **489** results.
20. The "Title" and "description" columns of this excel spreadsheet were collated with the other databases' Excel spreadsheets to begin the Screening Process document.

Scopus

- Interface/Platform: Scopus
- Date(s) accessed: 16/09/2024; 17/09/2024; 3/10/2024
- Who accessing: Gina Powell
- URL of source: https://www.scopus.com/
- Search Process:
    1. Selected "Advanced document search"
    2. Typed: "(child* AND behaviour AND intervention AND effect AND random*) AND (externalising OR disruptive OR ADHD OR "oppositional defiant" OR "conduct disorder" OR "intermittent explosive" OR antisocial) AND NOT meta-analy* AND NOT systematic AND NOT medic* AND NOT dose AND NOT stimulant AND NOT preschool AND NOT infant AND NOT kindergarten AND NOT internalising AND NOT anxiety AND NOT depression AND NOT telehealth LANGUAGE(english) DOCTYPE(ar)"



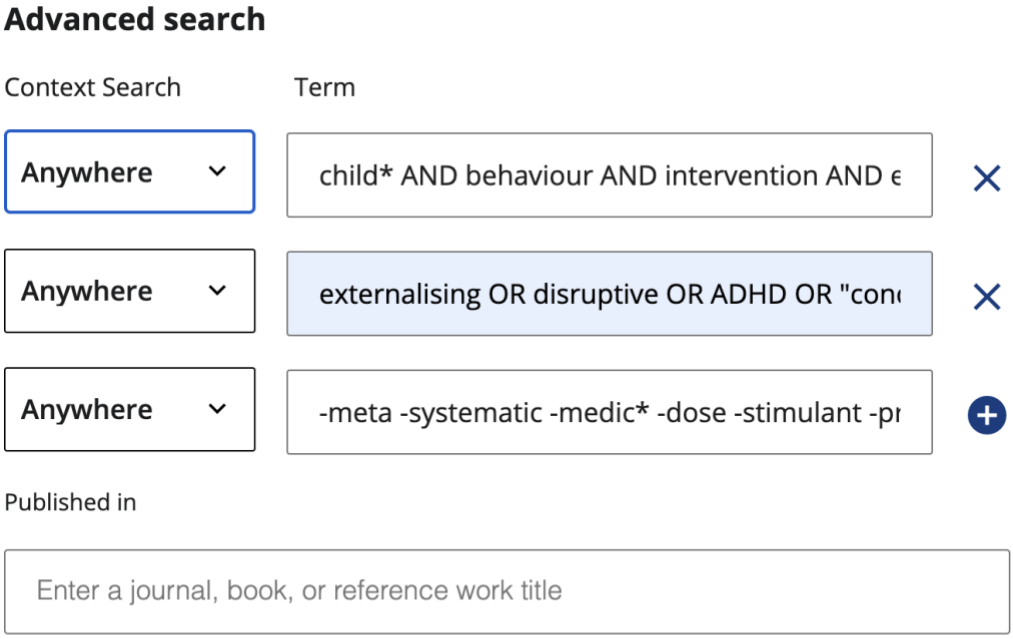
3. This search found 601 results before applying filters.

4. Under ‘Subject Area’ filter I limited to:
    i. Psychology
    ii. Social Sciences
    iii. Medicine
    iv. Arts and Humanities
    v. Neuroscience
    vi. Nursing
    vii. Health Professions
    viii. Decisions Sciences
    ix. Multidisciplinary
5. Under “Document Type” filter I limited to: “Article”.
6. Under “Language” filter I limited to: “English”.
7. Under “Keyword Filter” I excluded:
    i. Addiction
    ii. Adult
    iii. Alcohol
    iv. Alcohol Consumption
    v. Alcohol Drinking
    vi. Alcohol Use
    vii. Alcoholism
    viii. Cannabis
    ix. Drinking Behavior
    x. Middle Aged
    xi. Sexual Behavior
    xii. Substance Use
    xiii. Substance-Related Disorders
    xiv. Young Adult
8. Under “Source Title” filter I excluded:
    i. Addicta The Turkish Journal On Addictions
    ii. Canadian Journal Of Law And Society
    iii. Career Development For Exceptional Individuals
    iv. Chinese Journal Of Tissue Engineering Research
    v. Education And Training In Autism And Developmental Disabilities
    vi. Housing Policy Debate
    vii. Information Technology And People
    viii. International Journal of Consumer Studies
    ix. International Journal Of Educational Technology In Higher Education
    x. International Journal Of Social Robotics
    xi. Journal of Addictions Nursing
    xii. Journal of Child And Adolescent Substance Abuse
    xiii. Journal of Child Sexual Abuse
    xiv. Law and Human Behavior
9. Under “Publication Stage” filter I selected “Final”.
10. Under “Source Type” filter I selected “Journal”.
11. This search found **350** results.
12. All results were exported as a CVS file to excel.
13. The data was screened for duplicates in Excel by selecting the “Title” column, then selecting the Data tab and “Remove Duplicates” and “Expand the selection”, then deselecting “Select All” and reselecting “Title”. No duplicates were found.
14. **I returned to the saved search in Scopus and substited the search term “intervention” for “treatment”.**
15. This search found 496 results before applying filters.
16. Under ‘Subject Area’ filter I limited to:
    i. Psychology
    ii. Social Sciences
    iii. Medicine
    iv. Arts and Humanities
    v. Neuroscience
    vi. Nursing
    vii. Health Professions
    viii. Decisions Sciences
    ix. Multidisciplinary
17. Under “Document Type” filter I limited to: “Article”.
18. Under “Language” filter I limited to: “English”.
19. Under “Keyword Filter” I excluded:
    i. Addiction

   ii. Adult
   iii. Middle Aged
   iv. Substance Use
   v. Substance-Related Disorders
   vi. Young Adult
20. Under “Source Title” filter I excluded:
   i. Addicta The Turkish Journal On Addictions
   ii. Addictive Behaviours
   iii. Adicciones
   iv. American Journal On Addictions
   v. Applied Research On English Language
   vi. Canadian Journal Of Law And Society
   vii. Chinese Journal Of Tissue Engineering Research
   viii. Drug And Alcohol Dependence
   ix. Education And Training In Autism And Developmental Disabilities
   x. International Journal Of Engineering Pedagog
   xi. International Journal Of Social Robotics
   xii. Journal of Addictions Nursing
   xiii. Journal of Child And Adolescent Substance Abuse
   xiv. Journal of Child Sexual Abuse
   xv. Psychology Of Addictive Behaviors
21. Under “Publication Stage” filter I selected “Final”.
22. Under “Source Type” filter I selected “Journal”.
23. This search found **283** results.
24. I exported these results into excel, screened them for duplicates again. No duplicates were found.
25. I collated these 283 results to the other 350 and screened for duplicates again. 213 duplicate values were found leaving **420** results.
26. The “Title” and “description” columns of this excel spreadsheet were collated with the other databases’ Excel spreadsheets to begin the Screening Process document.

Wiley

- Interface/Platform: Wiley
- Date(s) accessed: 17th September 2024; 18th September 2024
- Who accessing: Gina Powell
- URL of source: https://onlinelibrary.wiley.com/
- Search Process:
   1. Selected “Advanced Search”.
   2. Search terms used were: (child* AND behaviour AND intervention AND effect AND random*); (externalising OR disruptive OR ADHD OR "conduct disorder" OR "oppositional defiant" OR "intermittent explosive" OR antisocial); (-meta-analys* -systematic -medic* -dose -stimulant -preschool -kindergarten -early -infant -internalising -anxiety -depression -telehealth).



3. This search found 572 results before applying filters.
4. Under “Publication Type” filter I selected “Journal”.
5. Under “Subjects” filter I selected ‘Psychology’.
6. This search found 227 results.
7. All results were exported in groups of 20 into Endnote then collated. 1 duplicate was found within Endnote which was deleted leaving 226 results.
8. The remaining 226 results were exported as an xml file into Excel. The change in file type duplicated some search results within Excel.
9. The data was screened for duplicates in Excel by selecting the “Title” column, then selecting the Data tab and “Remove Duplicates” and “Expand the selection”, then deselecting “Select All” and reselecting “Title”. 441 duplicates were found leaving **226** results.
10. **I returned to the saved search in Wiley and substited the search term “intervention” for “treatment”.**
11. This search found 439 results before applying filters.
12. Under “Publication Type” filter I selected “Journal”.
13. Under “Subjects” filter I selected ‘Psychology’.
14. This search found 90 results.
15. All results were exported in groups of 20 into Endnote then collated.
16. These 90 results were exported as an xml file into Excel. The change in file type duplicated some search results within Excel. 142 duplicate values were found leaving the **90** results.
17. I collated these 90 results to the other 226 and screened for duplicates again. 40 duplicate values were found leaving **276** results.
18. The “Title” and “Abstract” columns of this Excel spreadsheet were collated with the other databases’ Excel spreadsheets to begin the Screening Process document.

Appendix C

Outcomes Mapped to SWiM Constructs within Behavioural and Relational Domains

ID	Intervention	Outcome	Rater	Outcome Certainty
Behavioural				
Academic Functioning				
Active Engagement and Organisational Skills (Improve)				
52	EMR	Academic Engaged Time	Trained Observer	Moderate
92	PLAY	On-Task Behaviours	Trained Observer	High
128	DRC+	Academic Productivity	Teacher-Rated	Very Low
205a	GBGO	Academic Engagement	Teacher-Rated	Very Low
205b	PATHS+	Academic Engagement	Teacher-Rated	Very Low
232	EEST	Inhibition Executive Function Daily Behaviours	Parent-Rated	Very Low
*232	EEST	Initiation Executive Function Daily Behaviours	Parent-Rated	Very Low
232	EEST	Monitoring Executive Function Daily Behaviours	Parent-Rated	Very Low
232	EEST	Organizing Executive Function Daily Behaviours	Parent-Rated	Very Low
232	EEST	Planning Executive Function Daily Behaviours	Parent-Rated	Very Low
232	EEST	Shifting Executive Function Daily Behaviours	Parent-Rated	Very Low
232	EEST	Working Memory Executive Function Daily Behaviours	Parent-Rated	Very Low
257	STAND	Organisation (of School Materials)	Objective Coding	Moderate
*257	STAND	Planner Use (Academic Engagement/Organisation)	Objective Coding	High
861	GREET	Academic Engaged Time	Trained Observer	Moderate
988	RATIO	Academic Engaged Time	Trained Observer	Moderate
Class Disruptive Behaviours (Reduce)				
27	STARS	Disruptive Behaviour	Teacher-Rated	Very Low
44	FIRST	Maladaptive Behaviour	Teacher-Rated	Very Low
52	EMR	Disruptive Behaviour	Trained Observer	Moderate
92	PLAY	Disruptive Behaviours	Trained Observer	High
95	GBG	Disruptive Behaviour	Teacher-Rated	Very Low
128	DRC+	Deportment Impairment (Rule-Inappropriate Social-Behavioural Impairment)	Teacher-Rated	Very Low
861	GREET	Disruptive Behaviour	Trained Observer	Moderate
988	RATIO	Disruptive Behaviour	Trained Observer	Moderate

Engagement and Organisation Deficits (Reduce)				
89	MAGIC	ADHD-Related Executive Function	Parent-Rated	Very Low
*232	EEST	Learning and School Impairment from Behavioural Difficulties	Parent-Rated	Very Low
232	EEST	Life Skill Impairment (Managing and Organising Self)	Parent-Rated	Very Low
257	STAND	Academic Functional Impairment	Parent-Rated	Very Low
*257	STAND	Academic Functional Impairment	Teacher-Rated	Very Low
257	STAND	ADHD-Related Academic Problems	Parent-Rated	Very Low
257	STAND	ADHD-Related Academic Problems	Teacher-Rated	Very Low
Neurocognitive Functioning				
Attentional Dysregulation (Reduce)				
44	FIRST	Inattention	Teacher-Rated	Very Low
89	MAGIC	ADHD-Related Inattention	Parent-Rated	Very Low
95	GBG	Concentration Problems	Teacher-Rated	Very Low
128	DRC+	Attention Impairment	Teacher-Rated	Very Low
*166	PSYCHED	Inattentive	Teacher-Rated	Very Low
166	PSYCHED	Inattentive	Parent-Rated	Very Low
232	EEST	Inattention	Parent-Rated	Very Low
257	STAND	DBD Symptom Inattention Severity	Parent-Rated	Very Low
*257	STAND	DBD Symptom Inattention Severity	Teacher-Rated	Very Low
Disinhibition (Reduce)				
23	PR	Hyperactivity	Objective Device	High
44	FIRST	Hyperactivity	Teacher-Rated	Very Low
89	MAGIC	ADHD-Related Hyperactivity/Impulsivity	Parent-Rated	Very Low
128	DRC+	Academic Impulse Control	Teacher-Rated	Very Low
*166	PSYCHED	Hyperactive-Impulsive	Teacher-Rated	Very Low
166	PSYCHED	Hyperactive-Impulsive	Parent-Rated	Very Low
205a	GBGO	Hyperactivity	Teacher-Rated	Very Low
205b	PATHS+	Hyperactivity	Teacher-Rated	Very Low
*232	EEST	Hyperactivity-Impulsivity	Parent-Rated	Very Low
232	EEST	Risky Activities (Impulsivity and High-Risk Behaviours)	Parent-Rated	Very Low
257	STAND	DBD Symptom Hyperactivity/Impulsivity Severity	Parent-Rated	Very Low

*257	STAND	DBD Symptom Hyperactivity/Impulsivity Severity	Teacher-Rated	Very Low
Executive Dysfunction (Reduce)				
23	PR	ADHD Symptoms	Trained Observer	Very Low
43a	TALIP	Hyperactivity/Inattention (ADHD Combined)	Parent-Rated	Very Low
*43a	TALIP	Hyperactivity/Inattention (ADHD Combined)	Teacher-Rated	Moderate
43b	TALIB	Hyperactivity/Inattention (ADHD Combined)	Parent-Rated	Very Low
*43b	TALIB	Hyperactivity/Inattention (ADHD Combined)	Teacher-Rated	Very Low
107	MATE	ADHD-Related Behaviours	Teacher-Rated	Very Low
166	PSYCHED	ADHD Combined	Parent-Rated	Very Low
*166	PSYCHED	ADHD Combined	Teacher-Rated	Very Low
1015	FAST	Risk Development ADHD	Parent-Rated	Very Low
Emotional Awareness and Regulation (Improve)				
*107	MATE	Trait Emotional Intelligence	Child-Rated	Very Low
107	MATE	Trait Emotional Intelligence	Teacher-Rated	Very Low
205a	GBGO	Emotion Regulation	Teacher-Rated	Very Low
205b	PATHS+	Emotion Regulation	Teacher-Rated	Very Low
232	EEST	Emotion Control Executive Function Daily Behaviours	Parent-Rated	Very Low
Social Functioning				
Social Competence (Improve)				
27	STARS	Social Competence	Teacher-Rated	Very Low
44	FIRST	Adaptive Behaviour	Teacher-Rated	Very Low
*44	FIRST	Social Skills	Teacher-Rated	Very Low
44	FIRST	Social Skills Home-Based	Parent-Rated	Very Low
95	GBG	Prosocial Behaviour	Teacher-Rated	Very Low
107	MATE	Prosocial Behaviour PSBQ	Child-Rated	Very Low
205a	GBGO	Social Competence	Teacher-Rated	Very Low
205b	PATHS+	Social Competence	Teacher-Rated	Very Low
828	SEAT	Prosocial Behaviour	Peer-Rated	High
Antisocial Aggressive Behaviour (Reduce)				
89	MAGIC	ADHD-Related Defiance/Aggression	Parent-Rated	Very Low
107	MATE	Antisocial Behaviour CBCL	Child-Rated	Very Low

*107	MATE	Antisocial Behaviour CBCL	Teacher-Rated	Very Low
107	MATE	Antisocial Behaviour PSBQ	Child-Rated	Very Low
205a	GBGO	Aggressive-Disruptive	Teacher-Rated	Very Low
205b	PATHS+	Aggressive-Disruptive	Teacher-Rated	Very Low
828	SEAT	Overt Aggression	Peer-Rated	High
1015	FAST	Risk Development CD	Child-Rated	Very Low
*1015	FAST	Risk Development CD	Parent-Rated	Very Low
1132	ESSGP	Aggression	Peer-Rated	Very Low
Defiance (Reduce)				
27	STARS	Authority Defiance	Teacher-Rated	Very Low
44	FIRST	ODD Symptoms	Teacher-Rated	Very Low
257	STAND	DBD Symptom ODD Severity	Teacher-Rated	Very Low
257	STAND	DBD Symptom ODD Severity	Parent-Rated	Very Low
1015	FAST	Risk Development ODD	Child-Rated	Very Low
*1015	FAST	Risk Development ODD	Parent-Rated	Very Low
1091	AVOID	Non-Compliance	Teacher-Rated	Very Low
Externalising Behaviours Composite (Reduce)				
89	MAGIC	Difficult Child	Parent-Rated	Very Low
*89	MAGIC	Intensity of Child Behaviour	Parent-Rated	Very Low
89	MAGIC	Problem Behaviours	Parent-Rated	Very Low
92	PLAY	Externalising Behaviour	Teacher-Rated	High
*171	SELF	Daily Externalising Behaviour	Trained Coder	High
171	SELF	Externalising Behaviour	Parent-Rated	Very Low
1132	ESSGP	Externalising Behaviour	Parent-Rated	Very Low
*1132	ESSGP	Externalising Behaviour	Teacher-Rated	Very Low
Relational				
Impact of Others				
Parent Factors (Indirect)				
89	MAGIC	Parental Distress	Parent-Rated	Very Low
107	MATE	Authoritarian Parenting	Child-Rated	Very Low
107	MATE	Liberal Parenting	Child-Rated	Very Low

*107	MATE	Supportive Parenting	Child-Rated	Very Low
*166	PSYCHED	Parent Attitude	Parent-Rated	Very Low
166	PSYCHED	Parent Behaviour	Parent-Rated	Very Low
166	PSYCHED	Parent Knowledge	Parent-Rated	Very Low
257	STAND	Caregiver Strain	Parent-Rated	Very Low
Peer Factors (Indirect)				
828	SEAT	Class-Wide Cohesion	Peer-Rated	High
828	SEAT	Class-Wide Conflict	Peer-Rated	High
*828	SEAT	Class-Wide Cooperation	Peer-Rated	High
828	SEAT	Class-Wide Isolation	Peer-Rated	High
Teacher Factors (Indirect)				
92	PLAY	Teacher Stress from the Teacher-Child Relationship	Teacher-Rated	High
94	ECM	Inadequate Teaching Strategies in Compliance Scenario	Trained Coders	Moderate
94	ECM	Inadequate Teaching Strategies in Off-Task Scenario	Trained Coders	Moderate
94	ECM	Proactive Teaching Strategies in Compliance Scenario	Trained Coders	Moderate
*94	ECM	Proactive Teaching Strategies in Off-Task Scenario	Trained Coders	Moderate
94	ECM	Reactive Teaching Strategies in Compliance Scenario	Trained Coders	Moderate
94	ECM	Reactive Teaching Strategies in Off-Task Scenario	Trained Coders	Very Low
94	ECM	Teacher Attitudes - Negative Beliefs Towards Children with Challenging Behaviour	Teacher-Rated	Very Low
94	ECM	Teacher Attitudes - Positive Emotions Towards Children with Challenging Behaviour	Teacher-Rated	Very Low
94	ECM	Teacher Attitudes - Positive Planned Behaviours Towards Children with Challenging Behaviour	Teacher-Rated	Very Low
94	ECM	Teacher Attitudes - Positive Reactions Towards Children with Challenging Behaviour	Teacher-Rated	Very Low
94	ECM	Teaching Style - Psychological Pressure (Enragement and Disappointment)	Teacher-Rated	Very Low
94	ECM	Teaching Style - Warmth & Support	Teacher-Rated	Very Low
*166	PSYCHED	Teacher Attitude	Teacher-Rated	Very Low
166	PSYCHED	Teacher Behaviour	Teacher-Rated	Very Low
166	PSYCHED	Teacher Knowledge	Teacher-Rated	Very Low
*208	KEY	Teacher Interaction Skills Emotional Support	Trained Observer	High
208	KEY	Teacher Interaction Skills Proximity (Affinity and Cooperation Felt by the Students)	Child-Rated	High
Relationship Quality				
Parent-Child Relationship (Direct)				

89	MAGIC	Parent-Child Dysfunctional Interaction	Parent-Rated	Very Low
*107	MATE	Maternal Relationship	Child-Rated	Very Low
107	MATE	Paternal Relationship	Child-Rated	Very Low
232	EEST	Family Functioning Impairment	Parent-Rated	Very Low
257	STAND	Conflict with Child	Parent-Rated	Very Low
*257	STAND	Conflict with Parent	Child-Rated	Very Low
Peer-Child Relationship (Direct)				
23	PR	Peer Acceptance	Peer-Rated	High
89	MAGIC	ADHD-Related Peer Relations	Parent-Rated	Very Low
205a	GBGO	Peer Relations	Teacher-Rated	Very Low
205b	PATHS+	Peer Relations	Teacher-Rated	Very Low
232	EEST	Social Activities Impairment (Impaired Social Participation and Functioning)	Parent-Rated	Very Low
*828	SEAT	Acceptance	Peer-Rated	High
828	SEAT	Rejection	Peer-Rated	High
Teacher-Child Relationship (Direct)				
*27	STARS	Student-Teacher Relationship	Child-Rated	Very Low
27	STARS	Student-Teacher Relationship	Teacher-Rated	Very Low
52	EMR	Student-Teacher Relationship Quality	Teacher-Rated	Moderate
*208	KEY	Student–Teacher Closeness	Teacher-Rated	Very Low
208	KEY	Student–Teacher Conflict	Teacher-Rated	Very Low

Appendix D

Information on Intervention Mechanisms

Mechanism	Mechanism Description	How Mechanism Affect Outcomes	Strengths	Supporting Evidence
Behaviour Management Training	Teaches positive reinforcement and defined, clear and consistent management techniques for parents or teachers.	Improves recognition of appropriate positive antecedent and consequential strategies. Creates a structured environment that reinforces prosocial behaviours and diverts from negative escalating cycles through predictable, non-punitive consequences.	Improves individual and collective clarity on expectations and consequences creating a safe and stable environment. Reduces stress and improves confidence in behaviour management.	O'Connor & Hayes, (2020); Mabe et al., (2001)
Social Emotional Learning	Lessons teaching social and emotional literacy, self-awareness, regulation, reading social cues, healthy interactions, and empathy.	Develops understanding and skills for recognising and regulating emotions, promoting prosocial behaviours, and reducing reactive behaviour and conflict.	Reduces conflict, improves social and emotional intelligence, and interpersonal, academic and emotional outcomes.	Boncu et al., (2023); Cipriano et al., (2023)
Behaviour Tracking	Scaffolds and implements monitoring of target behaviours with reflection, review or feedback.	Regular reporting of targeted specific behaviours builds awareness, supports implementing reinforcement strategies, and promotes timely reflection and feedback to adjust interventions in real-time.	Promotes awareness, fidelity to treatments, accountability and adjustments or early intervention.	Iznardo et al., (2020)
Psychoeducation	Structured education about specific conditions or disorders, disorders and relevant coping strategies.	Improves recognition, understanding and attitudes toward disorders and related behaviour patterns, reducing misattributions and increasing consistent, and appropriate responses, which disrupt escalating cycles.	Supports positive outcomes through Empowers with understanding, promotes collective responsibility and reduces stigma.	Lukens & McFarlane, (2004)
Multisystem Collaboration	Ecological system collaboration between home, school, and other services to support better outcomes for the child.	Enables consistent goals, planning and expectations across contexts, and supports the identifying and implementing where support is needed across multiple contexts.	Shares the burden of responsibility across multiple parties, promotes open communication for consistent implementation of interventions.	Curtis et al., (2004); Walter & Petr, (2011)
Self-Management Skills Training	Practical skills training on goal-setting, time-management, planning, and organisation.	Improves ability to organise their resources, plan their time, set goals, and follow plans, providing a practical positive focus to reduce unfocused off-task behaviours.	Supports long-term confidence, a sense of achievement, autonomy and classroom success.	Bikic et al., (2017)
Attention Training	Digital cognitive training targeting differing aspects of attention through repetitive gamified skill practice.	Strengthens selective, sustained, and orienting attention, which supports executive skills, positive engagement with tasks and routine transitions.	Trains cognitive deficits associated with disruptive behaviours.	Kirk et al., (2017); Peng & Miller,

Play Therapy	Non-directive or structured play utilised to explore and resolve social or emotional challenges.	Provides a safe non-confronting space to explore emotions or social challenges through play that might usually be met with oppositional or aggressive behaviour.	Provides a space that allows social and emotional processing and learning to occur for children who might otherwise find it confronting.	Bratton et al., (2020); Meany-Walen, (2020)
Physical Environment Alterations	Altering the classroom or home physical environments to reduce various triggers.	Changes to routines or spatial and sensory factors to lower overstimulation, or arousal to provide an environment that supports regulation and appropriate responses.	Inclusive in nature, considers altering sensory/stimulating factors that decrease the need for behavioural intervention.	Barrett et al., (2015); Kinnealey et al., (2012); Wannarka & Ruhl, (2008)12/03/2026 7:25:00 pm
Learning Support	Differentiated instruction or extra tutoring to accommodate learning needs or difficulties.	Identifies and implements extra learning supports needed in turn reducing frustration and disengagement related to learning difficulties, lowers avoidance behaviours and promotes engagement.	Empowers children to engage in educational tasks by identifying and supporting their needs.	McIntosh et al., (2006); Morgan et al., (2008)