Unrecognised and Untreated: Improving the Identification of Mental Health Problems in Preschool-age Children through Examining Screening Measures, Multi-informant Reports and Symptom Stability

Rebecca Kar-Ling McLean

School of Psychology, Faculty of Science

University of Sydney

2025

A thesis submitted in fulfilment of the requirements for

the degree of Doctor of Philosophy

Acknowledgements

Mark Dadds, over the years you have always encouraged me to undertake the best science and to never limit my vision of what can be achieved. And you have empowered me to achieve that vision through your unwavering support. You ensured that the ideas were adequately resourced and fully developed. I am so grateful. I am particularly grateful that you managed to hold off the pull of vintage cars and blocks of timber for a little while longer. Enjoy this next stage of life; you have much to be proud of after such a distinguished career.

Lucy Tully, you have been a consummate supervisor and I could not have asked for more. At every stage of the project, you have been critically engaged with the ideas, method and papers. I have appreciated your thorough eye for detail, incisive questions and considered feedback. Beyond this research project, you have encouraged me to think more broadly about research career trajectories and have consistently acted as both a mentor and collaborator. I look forward to continuing those conversations and future projects with you.

Mark and Lucy, together you have been the dream supervisory team–responsive, supportive, dedicated. Jaimie Northam, Talia Carl and David Hawes, thank you for your insights and critical feedback on various project ideas, grant proposals and presentations. I particularly appreciated your input at the design stage of this project and re-design stage mid-pivot. Jaimie, you have been the ultimate champion, cheering and encouraging along the way. Thank you for your frank advice and support.

To all the clinicians and interns who worked on this project, thank you for your investment of time and effort into the many hours of clinical assessment. I especially want to thank Elysha Clark-Whitney for your clinical skills and being a sounding board in the final year of the PhD. Amongst the moments of frustration and uncertainty, it has been re-assuring to go on this ride together and thrilling to finish!

My thanks to the Child Behaviour Research Clinic & Growing Minds Australia teams (Tom, Dayna, Adrienne, Erika, Grace, Bridie, Olivia, Lindsay, Alex, Vilas, Jess, Lena, Tiana, Sharni, Torunn, Bronte) for the presentation feedback, questions, celebratory drinks, the rants and commiserations, and very importantly, the random bits of syntax… you are all an important part of a special team producing excellent research.

Finally, my heartfelt thanks to my family who has encouraged me throughout this long journey. Thank you for the sacrifices you have made to allow me to pursue an education and my ambitions in psychology. To the infant and preschooler and, now, preschool- and school-age children in my life, your learning, patience and boundless energy continue to surprise me. Thank you for the light-heartedness and joy you bring each day. To Dr Nahum "no notes" McLean, your efforts have not gone un-noticed and your achievements in the academic and personal space set the bar every day. Thank you.

Author Declarations

Statement of Originality

This is to certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose.

I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been acknowledged.

Rebecca K. McLean

Author Attribution Statement

Chapter 2 of this thesis has been published as McLean, R. K., Tully, L. A., Brinley, S. K., Carl, T., Turnell, A., Northam, J. C., & Dadds, M. R. (2024). Multi-Informant Universal Mental Health Screening for Preschool-aged Children by Parents and Educators: A PRISMA Systematic Review. *Clinical Child and Family Psychology Review*, *28*, 1–21. https://doi.org/10.1007/s10567-024-00506-2. I designed the study, developed the search strategy and performed the search, screened and extracted the data, conducted risk of bias assessment, analysed and interpreted the data and drafted the manuscript. L. Tully was involved in the conception and the design of the review, screened and extracted the data, conducted risk of bias assessment, and reviewed the manuscript. S. Brinley screened and extracted the data, conducted risk of bias assessment, and drafted parts of the results including tables and figures, and supplementary files. T. Carl screened and extracted the data, and conducted risk of bias assessment. A. Turnell screened and extracted the data, and conducted risk of bias assessment. J. Northam screened and extracted the data, and conducted risk of bias assessment. M. Dadds was involved in the conception and the design of the review, and reviewed the manuscript. All

authors contributed to the development of the research questions and selection criteria; read, reviewed and approved the final manuscript.

Chapter 3 of this thesis has been published as McLean, R. K., Tully, L. A., & Dadds, M. (2025). Reliability, predictive validity and normative data for the Pediatric Symptom Checklist-17 (PSC-17) in a national Australian sample. *Australian & New Zealand Journal of Psychiatry*, *59*(8), 702–712. https://doi.org/https://doi.org/10.1177/00048674251342952. I designed the study, analysed and interpreted the data, and drafted the manuscript. L. Tully was involved in the conception and the design of the study, reviewed the data and the manuscript. M. Dadds was involved in the conception and the design of the study, reviewed the data and the manuscript. All authors contributed to the development of the research questions; read, reviewed and approved the final manuscript.

Chapter 4 of this thesis has been published as McLean, R. K., Tully, L. A., & Dadds, M. R. (2025). Multi-informant reports of preschool mental health: Validation of parent and educator reports and normative data for the preschool Pediatric Symptom Checklist and PSC-17. *Child and Adolescent Psychiatry and Mental Health*, *19*(1), 138. https://doi.org/10.1186/s13034-025-00985-3. I designed the study, analysed and interpreted the data and drafted the manuscript. L. Tully was involved in the conception and the design of the study, reviewed the data and the manuscript. M. Dadds was involved in the conception and the design of the study, reviewed the data and the manuscript. All authors contributed to the development of the research questions; read, reviewed and approved the final manuscript.

Chapter 5 of this thesis has been submitted for publication. I designed the study, analysed and interpreted the data and drafted the manuscript. L. Tully was involved in the conception and the design of the study, reviewed the data and the manuscript. M. Dadds was involved in the conception and the design of the study, reviewed the data and the manuscript.

All authors contributed to the development of the research questions; read, reviewed and approved the final manuscript.

For the studies reported in Chapters 3 and 4a, Ms Domna Alloush assisted with data preparation by cleaning the data and conducted comparisons of the sample to Australian census data.

During the preparation of the thesis, I engaged Dr Christina Houen of Perfect Words Editing for the purposes of copyediting the thesis. I confirm that where text was modified, the content was reviewed, and I take full responsibility for the submitted thesis.

Rebecca K. McLean

26 August 2025

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Mark R. Dadds

26 August 2025

Lucy A. Tully

26 August 2025

Artificial Intelligence

No content produced by generative AI tools has been used in the preparation of this thesis.

Funding Statement

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

The research in Chapters 3–5 was supported by Growing Minds Australia: A National Trials Strategy to Transform Child and Youth Mental Health Services, and the Growing Minds Australia Clinical Trials Network (Medical Research Future Fund: MRF2006438).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CBCL	Child Behavior Checklist
CGAS	Children's Global Assessment Scale
DISCAP	Diagnostic Interview Schedule for Children, Adolescents and Parents
ECEC	Early Childhood Education and Care
ICC	Intraclass Correlations
MH	Mental Health
NPV	Negative Predictive Value
Preschool PSC	Preschool Pediatric Symptom Checklist
PPSC	Preschool Pediatric Symptom Checklist
PPV	Positive Predictive Value
PSC	Pediatric Symptom Checklist series of measures
PSC-17	Pediatric Symptom Checklist-17
SDQ	Strengths and Difficulties Questionnaire
SEB	Socio-emotional and Behavioural
UMHS	Universal Mental Health Screening

Abstract

Despite evidence for the high prevalence of mental health (MH) problems in children aged 3–5 years, parents of preschool-age children are less likely to recognise and access support for MH problems compared to parents of older children. This thesis evaluates screening measures for young children, which could help parents and educators identify children at risk of MH problems. Chapter 2 presents a systematic review of existing multi-informant MH measures for young children and examines their predictive and incremental validity, effectiveness and acceptability. Chapters 3–4 present a series of studies evaluating the reliability, validity and acceptability of two versions of the Pediatric Symptom Checklist (PSC), the Preschool PSC and PSC-17, reported by parents and educators, in a national, cross-sectional study. Chapters 4–5 present a longitudinal study in which a paired, community sample of parents and educators were tested at three time points to examine incremental and diagnostic validity, and symptom stability.

The systematic review concluded that several existing measures had acceptable predictive validity, but there were many shortcomings concerning incremental validity (Chapter 2). Subsequent studies reported strong psychometric properties for both PSC measures, including internal consistency, concurrent validity, and high acceptability (Chapters 3–4). Parent reports improved the prediction of functioning scores over and above educator reports supporting incremental validity. PSC measures had moderate–strong associations with diagnostic severity and functioning scores (Chapter 5). Total scores for both PSC measures, PSC-17 externalising and attention subscales had significant measurement stability across time. Together these studies demonstrate the validity, acceptability and clinical utility of the PSC measures, and contribute new evidence supporting measures that can be used for multi-informant, universal screening of MH problems in preschool children.

CHAPTER 1: INTRODUCTION

Despite the early onset, high prevalence and prognostic value of socio-emotional and behavioural or mental health (MH) problems in preschool-age children 3–5 years, parents of young children are less likely to recognise emerging difficulty and less likely to access services and support for child MH concerns compared to parents of older children (Huang et al., 2019; Segal et al., 2018). Without early identification and resulting access to early intervention, MH problems can become more severe, persist into adolescence and adulthood, and have long-term negative ramifications across health and education domains (Jones et al., 2015; Tremblay et al., 2004). Therefore, the aim of this thesis was to evaluate and validate brief screening measures to help parents and educators identify preschool children at risk of MH problems, in order to improve the likelihood that such children receive timely, early intervention.

1.1 Unrecognised and Untreated Mental Health Problems in Preschool-age Children

The age of compulsory schooling in Australia is six years, which may be the first time that a child is reviewed for cognitive, behavioural, emotional or developmental challenges. However, children are likely to have emerging MH problems in the years preceding school entry (Campbell et al., 2000; Côté et al., 2009). Prior to school entry, children aged 3–5 years experience a period of rapid development, and it can be difficult for parents or primary caregivers (hereafter referred to as parents) to discern whether emotions and behaviours displayed during this period require support and intervention. Rapidly changing behaviour and social and emotional development during this period can make accurate appraisals of child wellbeing difficult for parents, particularly when behaviour is transient and varies across contexts and environments (Konold et al., 2004). Aggressive, non-compliant, impulsive, hyperactive behaviour, fears and worries in young children can be age appropriate, relatively normative and transient; however, if severity or impact is high, these behaviours can also be

symptomatic of emerging MH problems (Campbell et al., 2000). Differentiating between transient, normative behaviour and more stable, emerging trajectories of psychopathology is a challenge for parents of preschool-age children (Sim, Thompson, Marryat, Ramparsad, et al., 2019). For example, a representative survey of Australian parents found that 57% of parents did not know that it was not normal for preschool-age children to be consistently disobedient (Rhodes, 2017). Although socio-emotional and behavioural symptoms can be detected in children as young as 18 months–2 years (Bayer et al., 2012; Tremblay et al., 2004), parents of young children frequently struggle to identify children who are in need of additional support. Parents of younger children (aged 4–7 years) in particular are significantly less likely to recognise MH problems compared to parents of older children (8–15 years), regardless of household income (Huang et al., 2019). Recent research has also shown that parents of children aged 3–5 years are three times less likely to recognise MH problems and two times less likely to perceive a need for help for MH problems than are those of older children aged 12–17 years (Alloush et al., 2025).

A large body of work has established that many MH problems first emerge in children aged 3–5 years. MH problems in preschool-age children can include internalising problems such as fears, worries, anxious and depressive symptoms, and externalising problems such as defiance and hyperactive and aggressive symptoms (Achenbach & Edelbrock, 1978). Several reviews have indicated that problems are present in community and clinical populations, but prevalence rates in preschool-age children vary internationally, even amongst predominantly Western, educated, industrialised, rich, and democratic (WEIRD) nations. For example, in the United States (US), a review of four studies found 14–26.4% of children aged 2–5 years have some psychopathology or psychiatric disorder (Egger & Angold, 2006), whilst three studies including community samples of children aged 4–7 years from the United Kingdom, Iceland and Norway reported lower prevalence rates of 7.1–10.1% (Gudmundsson et al., 2013). All

studies utilised comprehensive questionnaire and structured diagnostic interviews to assess children. A large Australian cohort study found 11.5% of 2-year-old and 20% of 5-year-old children had clinically significant behavioural problems (Robinson et al., 2008) and another Australian study found 13–23% of children aged 1.5–6 years had elevated symptoms of MH problems (Oh et al., 2015). Both studies assessed child MH problems using the Child Behavior Checklist (CBCL; Achenbach, 1999), a comprehensive parent-report questionnaire. Taken together, the prevalence estimates among preschool-age children range widely from 7.1% to 26.4%, with differences likely due to varying methodology and ages of included children. However, it is clear that the preschool years are a critical period for the emergence of MH problems, and represents an important window of opportunity for early intervention before problems become more entrenched and severe. Early intervention programs are designed to support a child's development and wellbeing early in the trajectory of the identified problem. Programs can include a broad array of supports and are delivered to children deemed at-risk of developing problems (Ramey & Ramey, 1998). Connecting at-risk children to early intervention and support services, however, can be challenging.

Whilst the emergence of MH difficulties in young children has been well established, few children receive adequate support for these difficulties and rates of unmet needs are concerning. In Australia, fewer than one in three children or adolescents with MH difficulties access and receive support from health professionals (Hiscock et al., 2020). Amongst children aged 8–13 years who met clinical thresholds for internalising or externalising symptoms, 90.9% had not attended a single MH appointment. Within the preschool-age range, less than one sixth (15%) of Australian children with elevated MH symptoms received any type of professional healthcare (Oh et al., 2015), and only 3% of 4-year-old children with diagnosed clinical disorders in the US received MH services (Lavigne et al., 2009). Not only are younger children less likely to receive healthcare services, but they have the lowest rates of access to

specialist MH services (Segal et al., 2018). Young children are severely underserviced relative to older children; this is an important gap, given that early intervention is likely to be more effective and cost-effective compared to later treatment.

Unrecognised and untreated, children with MH difficulties can experience far-reaching negative sequelae throughout adolescence and into adulthood. The long-term ramifications of untreated and unmet MH needs span domains such as employment, education, criminal activity, mental and physical health throughout later childhood and adulthood (Jones et al., 2015). In the preschool-age range, children who do not learn to regulate aggression prior to school entry are at high risk of serious, more aggressive behaviour in later adolescence and adulthood (Tremblay et al., 2004). The early age of onset of mental disorders is associated, not only with longer periods of untreated illness, but also with poorer clinical and functional outcomes (de Girolamo et al., 2012). This highlights the importance of helping parents to recognise and identify MH concerns in young children as the first step towards receiving early intervention. However, parental attitudes, beliefs and lack of knowledge about problematic behaviour can hinder recognition of child MH problems. MH problems in preschool-age children are often left untreated, since parents normalise MH difficulties as typical challenging behaviour in this age, believing problems will resolve on their own (Rhodes, 2017). Not only do parents believe children will grow out of their problems, but they are also not aware of the benefits of seeking early intervention despite the established evidence base (Alexander et al., 2013; Pavuluri et al., 1996). In order to combat these beliefs, parents need measures which will help them recognise child MH problems so that they can take subsequent steps to seek early intervention for their child, thereby altering the trajectory of MH problems.

1.2 Universal Mental Health Screening

Identifying children who are at risk of MH problems enables children to be directed towards early intervention and potentially halts the negative prognosis of untreated problems. This identification process can occur through universal screening, which is also known as systematic monitoring, routine surveillance and wellbeing check-ins. Universal MH screening is a population-based approach in which systematic assessments or screening measures are used to identify children who have MH problems. Unlike traditional medical models in which screening refers to the detection of a disease prior to the manifestation of symptoms (Eddy, 1985), screening in a MH context identifies children with elevated symptoms who may require further assessment and may benefit from intervention. It encompasses the examination of all children in a specified population and can comprise single or multiple informants; single assessments or multistage, multi-gate assessments which triage children for further assessment. Although some researchers contend that non-standardised assessment can form part of screening, the findings of several studies have brought into question the utility of screening programs which incorporate non-standardised assessments such as paediatric assessment or teacher nomination methods (Humphrey & Wigelsworth, 2016; Sheldrick et al., 2011; von der Embse & De Los Reyes, 2024). In fact, direct comparisons have found that teacher nomination identified less than half the number of at-risk students (aged 6–13 years) as identified by standardised, teacher-report screening measures (Eklund & Dowdy, 2014). Primary characteristics of universal screening include standardised assessments that include a clinical threshold or cut-off score to indicate whether symptoms are severe enough to warrant further investigation. Once identified as at risk, children can be triaged for further assessment and then referred to appropriate early intervention as required (Moore, Dowdy, Hinton, et al., 2022). Screening is time-efficient for users when measures are brief and is a cost-effective mechanism

for overburdened healthcare systems faced with issues of accessibility and long waitlists (Humphrey & Wigelsworth, 2016).

Currently, the preschool period represents a gap in which routine screening is missed for lengthy periods. Typically children in the first two years of life attend multiple immunisation, developmental milestone and wellbeing health checks from birth to 18 months (Australian Government, 2024), and whilst many Australian health jurisdictions recommend routine health monitoring annually between 2–4 years, in practice up to 20% of children have not received complete vaccination schedules despite parent beliefs that they are complete (Lim et al., 2019). Since MH and developmental monitoring are tied to vaccinations, this means that a substantial proportion of children may not receive attention for MH and wellbeing until school entry at six years. Thus, there are several years in which MH problems may emerge and children may not receive adequate early intervention.

The success of early interventions for MH problems in preschool children is dependent on the early detection of children who could benefit from targeted support; however, early detection of MH problems through universal screening is only common in a limited number of countries (Bayer et al., 2009). Standardised, universal, developmental screening is mandated in parts of the developed world such as Scotland, Denmark, Norway and the US, although there is little consensus on how to deliver such programmes (Wilson et al., 2018). Previously, in 2008, Australia adopted a Healthy Kids Check (HKC) which entailed the universal screening of all 4-year-old children to detect the early signs of physical and developmental concerns, and was conducted by general practitioners (Alexander et al., 2013). In 2011, a proposal to amend the screening to explicitly include social and emotional development questions and to decrease the age of the HKC from four to three years of age was heavily criticised in public media and ultimately abandoned (Dunlevy, 2012). Criticisms of the HKC included the potential risks of over-medicalisation, stigmatisation and false diagnosis of children, in addition to the potential

to induce parental anxiety about child MH (Dunlevy, 2012; Frances, 2012; Jureidini & Raven, 2012; Prior, 2012). The very limited published research on the HKC showed that these concerns did not have a strong evidence base. There was no evidence of iatrogenic effects, such as increased stigma or parental anxiety, since these outcomes were not specifically examined in an evaluation of two medical practices in Queensland, comprising only 557 cases out of a possible 282,200 4-year-old children across Australia at the time (Thomas et al., 2014). Nevertheless, the HKC program was ultimately defunded in 2015, in part due to poor acceptability of the screening by a range of stakeholders including politicians, general public, parents, and researchers from outside of Australia (Alexander & Mazza, 2015).

Acceptability, also known as social validity, encompasses whether screening is considered appropriate, helpful or wanted by those who undertake the screening such as parents, educators and clinicians (Humphrey & Wigelsworth, 2016). It can also ascertain if users find measures easy to comprehend and whether they perceive the questions to be important. Acceptability can be assessed for screening programs or specific screening measures. However, there is limited research exploring the acceptability of screening measures or programs, in spite of the World Health Organisation's long-standing stipulation that acceptability is a key principle of any screening program (Wilson & Jungner, 1968). This lack of research has resulted in the rejection of developmental screening programs being implemented in parts of the world such as the United Kingdom (UK National Screening Committee, 2005). A recent review of research on universal screening for children aged birth–18 years found that only seven studies across the world and across health and education settings reported on acceptability (Brinley et al., 2024). The review found parents perceived MH screening to be appropriate and helpful regardless of setting and that they may be willing to engage in screening initiatives (Brinley et al., 2024). A small review of four preschool screening measures found only one study investigated acceptability and reported the Ages and

Stages Questionnaires: Social-Emotional was acceptable to parents (Feeney-Kettler et al., 2010). If screening measures are not acceptable, then users such as parents and educators are less likely to utilise them, and the screening program will be ineffective.

The acceptability of universal screening (or lack thereof) can hamper efforts to implement screening across clinical and community settings, which highlights the importance of ascertaining the acceptability of screening, particularly amongst preschool parents and educators (Glover & Albers, 2007; Harrison et al., 2013; Kamphaus et al., 2007). Given the previous criticism of universal MH screening of young children in Australia and the scarcity of research in this area, acceptability is an essential construct to investigate if screening is to be widely adopted as a strategy to support access to early intervention. Thus, the acceptability of screening measures amongst parents of preschool-age children is an outcome of interest and remains a significant gap in the research which needs investigation.

1.3 Parents' Role in the Help-Seeking Process

As young children are unable to access healthcare and support services for their own wellbeing, parents and other caregivers play a pivotal role in seeking help for the child. Help-seeking models situate responsibility for the process on parents and other caregivers, such as educators, but place particular emphasis on the responsibility of the parent (Costello et al., 1998). Parents act as gatekeepers for the child's MH care by facilitating access to services and promoting help-seeking behaviour in older children and adolescents (Schnyder et al., 2020). Parents' access to support for their child has been theorised as the help-seeking process in which parents first, recognise the child has a problem; second, form the intention to seek professional help; and third, access or receive help to support their child's wellbeing—although the specific number of steps varies between models (Logan & King, 2001; Reardon et al., 2017). The initial behavioural model proposed by Andersen and Newman (Andersen, 1995)

suggests that there are a number of pre-disposing factors such as demographic, gender, or ethnic characteristics, and enabling factors such as parental education and family income, which precede a parent's perceived need for help. Factors such as parent age, ethnicity and previous help seeking have been associated with higher odds of help-seeking behaviour for MH problems in 3-year-old children (Luo et al., 2022). A common focus across models is the importance of parent perception or recognition of MH problems in their child. Parental recognition of a problem, along with perceived parental burden or functional impact of the child's problem, is what drives subsequent steps to access to MH services (Costello et al., 1998). Without the crucial first step of recognising MH difficulties, parents do not take the subsequent steps in the help-seeking process and therefore leave children with unrecognised and untreated problems (Reardon et al., 2017). Apart from lack of recognition by parents, children may not receive help because parents do not have confidence or knowledge of how to access support for child MH difficulties. A representative survey of Australian parents reported that almost two-thirds of parents (65%) lacked the confidence to recognise signs of MH problems in their child, and parents of young children often did not know where to seek help (Rhodes, 2017). This lack of confidence and knowledge can mean that help seeking is delayed as parents seek out various sources of help to find out more about their child's difficulties, or help may not be sought at all.

An important distinction in the discussion of parent help seeking is whether sources of help sought are formal or informal. Formal sources of help include professional health-service providers with specified roles in MH service delivery, such as psychologists and general practitioners (Rickwood et al., 2012). This category can be further delineated into specialist (e.g., psychologists, psychiatrists) and non-specialist formal roles such as general practitioners, paediatricians, and child and family health nurses (Oh et al., 2015). Informal sources of help include untrained sources of help such as family members, friends and individuals in the

community, who may, in ideal circumstances, connect parents with formal support services (Farrelly, 2008). Some researchers have added an additional category of help such as semi-formal sources, which include professionals without MH training such as educators or teachers (Rickwood et al., 2012). Formal sources of help such as healthcare professionals have been researched extensively and found to entail significant barriers to parental help seeking, such as limited accessibility and long waitlists. In addition, the financial cost of obtaining professional help is often prohibitive and has been rated by preschool parents as a major barrier to accessing help (Harwood et al., 2008). Parents who are unable or unwilling to access formal help frequently turn to alternative, informal sources for advice and validation of concerns. The informal or semi-formal help-seeking area is important to consider, since a large number of parents seek help from these sources in the first instance. A large Australian longitudinal study found 36–45% of parents sought help from informal sources, whilst only 7–8% accessed non-specialist, formal sources of care despite having children (aged 1.5–3 years) with elevated MH symptoms (Oh et al., 2015). Moreover, parents of preschool-age children initially seek help from informal sources prior to accessing formal sources of care, often seeking validation from these sources prior to presenting at health services (Alexander et al., 2015; Pavuluri et al., 1996). Importantly, more parents of preschool children seek help from preschool staff than specialists or general practitioners, a finding that has been repeated across studies (Oh et al., 2015; Pavuluri et al., 1996). Concerningly, 38% ($n = 42$) of mothers of preschool children have also indicated that they would *never* or *rarely* speak to a paediatrician about problem behaviour despite harbouring concerns (Harwood et al., 2008). Thus, educators may play a key role in the promotion of help seeking for parents of young children with emerging MH problems.

Educators' role in the help-seeking process may involve helping parents to recognise a child's MH difficulties and/or referring parents to formal sources of support. A key study demonstrates the potential impact that educators can have in the help-seeking process. In older

children, teacher recognition of child MH difficulties combined with parental perception of MH impairment or perceived burden increased access to MH services at a rate six times the average rate of access to school-based services (Costello et al., 1998). This suggests that parents may be open to acting on advice from educators or other healthcare professionals if concerns are raised with them. However, research has shown that healthcare professionals and educators tend to wait for parents to raise concerns about their child's wellbeing in the first instance (Alakortes et al., 2017). In fact, research shows that neither parents nor educators raise concerns about child MH with each other spontaneously, instead waiting for the other party to alert them to potential problems (Burkett et al., 2025; Honda et al., 2023). This suggests that there is a need for establishing routine, standardised screening so that both parents and educators have an opportunity to complete measures on child MH and discuss concerns.

1.4 The Role of Preschool Educators in Multi-informant Assessment

Multi-informant assessment is a hallmark of developmental research as it captures reports about a child from multiple perspectives and accounts for individual biases, knowledge and understanding of the child in specific contexts. It assesses child socio-emotional and behavioural problems across contexts, with pervasiveness indicating a more severe or persistent expression of MH disturbance (De Los Reyes et al., 2015). The expression of symptoms across contexts, such as home and education settings, can also be indicative of risk for long-term problems (Campbell et al., 2000). Reports from outside the home are particularly important for conditions such as attention-deficit/hyperactivity disorder, in which the expression of symptoms is highly context dependent (Campbell et al., 2000). Assessment in the form of universal screening that relies exclusively on single informants may omit important information about the child and may miss at-risk children (Eklund & Dowdy, 2014; Feeney-

Kettler et al., 2011). Therefore, reports from both parents and educators are important perspectives to capture about a child's MH.

Parents can provide important data about the child from the home and social context, whilst outside the home, educators are likely to be effective informants about preschool children in their care for several reasons. First, given the amount of time they spend with children, educators can develop familiarity with individual children's social skills, behaviours and emotions. In Australia, 4–5-year-old children attend early childhood education and care (ECEC) 15–21 hours per week on average (Australian Bureau of Statistics, 2021b). ECEC can include preschools, centre-based day care, or long day care. Second, educators observe children in naturalistic play with peers and have a referent group of children of a similar age with which to compare typical child behaviour and wellbeing (DeLucia et al., 2022). This is an advantage of educator perspectives that sets it apart from assessments taking place in clinical settings or home settings. Third, compared to formal MH services, educators and ECEC settings are less encumbered by issues of access and stigma, making them an ideal setting for universal screening (Desta et al., 2017). One study of children from preschool through to year one found most parents and teachers agreed that the preschool setting is appropriate for screening MH difficulties; less than 2% of parents (n = 7) and less than 10% (n = 4) of teachers disagreed with the importance of universal MH screening in the pre/school context (Moore, Dowdy, Hinton, et al., 2022). This finding is also reflected in studies with teachers from across ECEC and primary schools, with 89% (n = 260) agreeing that schools should be involved in addressing the MH issues of students (Reinke et al., 2011). However, the acceptability of screening of preschool-age children by both parents and educators has generally not been well researched. More research has been conducted in older age ranges, which has found teacher screening is considered acceptable to parents (Soneson et al., 2018). Finally, educators themselves have identified that they are willing to play a part in the prevention and

identification of MH problems in children and believe that preschools should have a role in addressing the MH needs of children (Eleni, 2021; Reinke et al., 2011). A recent qualitative study with Australian primary care clinicians found that clinicians believed there is an opportunity for clinicians and school teachers to work more collaboratively, with formal communication mechanisms established to support child MH together (Paton et al., 2022). Taken together, these findings suggest that educators may have an important role to play in the help-seeking process and supporting the recognition of MH problems. However, despite the potential value of educators' insight into child wellbeing, educators frequently lack formal training in child MH (Stormont & Stebbins, 2005). They may be willing to play a role in the screening and support of children with MH needs, but educators often lack training and important knowledge around evidence-based practices (Reinke et al., 2011). This then means there is a need to ensure educators are equipped with accurate screening measures that capture their insights into children's wellbeing.

1.5 Validation of Universal Mental Health Screening Measures

There are a plethora of child MH measures available, but few are suitable for multi-informant, universal screening in the preschool period. Such measures need to be brief, freely available, offer parent- and educator-report, suitable for use with large populations of preschool children aged 3–5 years, easy to score, validated and ideally, able to be administered and scored online. Measures can be tested against standards of reliability and validity in a process of measure validation, which can then determine whether a measure is suitable for use with the population with which it was validated. Typically, tests of reliability include internal consistency, test-retest reliability, and inter-rater reliability. Tests of validity can include criterion, content and construct validity. Although definitions and taxonomy of measurement

properties are debated (Mokkink et al., 2010), there are several measurement properties which are key to considerations of multi-informant, screening measures.

Incremental validity is a measurement property that is of particular importance when considering the use of multi-informant measures. It provides an indication of how much each informant's perspective contributes to the outcome and what their context-specific perspective adds to an understanding of the child's MH and wellbeing (De Los Reyes et al., 2015). Since multi-informant measures require more input from multiple caregivers in terms of time and burden, there is a need to ensure that this is justified. Adding together educator- and parent-reports of child MH may have the potential to increase the validity of measures, but this needs to be investigated in research. Currently, there is little research that has examined whether preschool educator ratings of child MH have incremental validity over and above that of parent ratings. Criterion validity is also an important consideration for universal screening, since it indicates the level at which a screening measure aligns with an established standard or outcome such as clinical diagnosis or elevated MH symptoms. Criterion validity encompasses the validation of the test measure against a criterion measure or other construct, which can be conducted at one timepoint (as in concurrent validity) or longitudinally in a comparison of a screening measure's outcome to a later diagnostic or classification outcome (as in predictive validity).

Predictive validity is particularly important in universal screening, since screening measures which are more accurate in identifying future outcomes have greater clinical utility. Predictive validity analysis can produce rates of sensitivity and specificity, which are rates of correctly identified positive and negative cases, that is, children who are or are not at risk of MH problems. Whilst screening measures are not diagnostic, it is essential that they accurately identify children at risk so that children can be appropriately referred for early intervention. When screening measures are not accurate, they can incorrectly categorise children as at risk

when they are not, or conversely, miss identifying at-risk children in need of intervention. Both under- and over-identification can lead to iatrogenic effects such as unwarranted stigmatisation of the child, parental anxiety, or criticism of informants, the measure or screening process. Moreover, inaccurate screening can lead to wasted resources by unnecessarily referring children to health practitioners or missing the opportunity to provide lower intensity, early intervention, thereby undermining the effectiveness of screening (Sawyer et al., 2014). Currently, there is a need for more research investigating the predictive validity of screening measures in preschool-age children using comprehensive questionnaire criterion measures or diagnostic clinical interviews.

Clinical thresholds are important to establish for screening measures as they provide a systematic classification of children above or below specified cut-off scores. Referral of children to early intervention depends on the identification of a clinical threshold or cut-off score which determines the classification of children as at risk or not at risk. Clinical thresholds determine the level of risk ascribed to children and therefore, the identification of risk in children is critical to the success of screening interventions (Moore, Dowdy, Fleury, et al., 2022; Sheldrick et al., 2015). Appropriate clinical thresholds can be ascertained by testing a measure in the population of interest and establishing normative data, in addition to tests of predictive validity**.** Normative data is the distribution of scores within a normative sample and provides a frame of reference for the interpretation of results (Mitrushina et al., 2005). A normative sample is typically a large, representative sample of the population of interest; in the case of universal screening, the target population is most often a general, non-clinical, community population. As such, requisite demographic criteria, such as restrictions on eligibility, should be as minimal as possible. However, this means that obtaining normative data can be costly when the composition of the normative group must be carefully prescribed and inclusive. The size of the normative sample must also be sufficiently large. However, what

constitutes a sufficient sample size is contentious and there are no set conventions regarding minimum sample sizes (Kendall & Sheldrick, 2000; Mitrushina et al., 2005). Nevertheless, normative data is an important component of both screening programs and measures, as it enables the systematic interpretation of results and efficient identification of children at risk of MH problems by specifying the threshold at which children should receive further attention (Sheldrick, Henson, Merchant, et al., 2012). For decision making about children who score in the at-risk range and may benefit from early intervention, screening measures must be validated in relevant populations and have verified clinical thresholds.

Clinical thresholds serve an additional purpose in terms of identifying children with elevated MH symptoms. The measurement and study of stability of MH problems can be challenging during the preschool period, since many disruptive behaviours are developmentally normal during this period. However, elevated MH symptoms, as indicated by scores on validated measures which exceed clinical thresholds, can indicate those at risk of continuing trajectories of MH problems (Sheldrick et al., 2025). The importance of identifying preschool-age children scoring above clinical thresholds on screening measures is indicated by research showing that elevated symptoms in childhood—more so than diagnosis—are associated with later adult MH disorder (Mulraney et al., 2021). This approach emphasises the utility of screening in childhood as important for later adult or adolescent MH outcomes, which may not have been identified in childhood. An alternative approach to conceptualising the use of clinical thresholds is decision analysis, a type of clinical decision making that includes diverse attributes of utility. Within this framework, thresholds are set based on various attributes, which are not limited by the psychometric performance of screening measures, but take into account the possible benefit and costs of referring a child for further investigation (Pauker & Kassirer, 1980). Thus, additional factors such as the availability of resources, and

potential burden of further investigation in the individual and collective context may be considered in decision making and setting of thresholds.

1.6 Need for Validated Mental Health Screening Measures in the Preschool Period

There have been no systematic reviews of multi-informant, universal screening measures focussed on preschool-age children to date. Given this gap, there is a need to identify the available measures for parents and educators in this age range and to examine their psychometric properties. A recent review of free, brief and accessible child MH measures by Becker-Haimes et al. (2020) identified only two measures with "excellent" psychometric properties which could be used with preschool-age children. The two measures identified were the Pediatric Symptom Checklist (PSC; Jellinek et al., 1988) and the Strengths and Difficulties Questionnaire (SDQ; Goodman & Goodman, 2009). Both measures are broad measures of child MH that are suitable for use as screening measures, but only one of these, the SDQ, has been validated for multi-informant use.

The SDQ is a 25-item questionnaire that produces a total difficulties score, and five subscales including conduct, hyperactivity, emotional, peer problems and prosocial skills. There are two versions which have been validated for children 2–4 years and 4–17 years. It is one of the most widely used measures of child MH across the world, including in Australia, across clinical and research settings (Vaz et al., 2016). The long history of use of this measure means that there is a large dataset upon which researchers can draw and a substantial body of research validating its properties, and yet numerous concerns have been raised about its psychometric properties. In a major review of 41 validation studies of the SDQ when used with children aged 3–5 years, there was limited support for discriminant validity, predictive validity, internal consistency for subscales and test-retest reliability (Kersten et al., 2016). A separate review of screening measures for children of all ages also reported that no studies reported

simultaneously high sensitivity and high specificity for the SDQ, which are key indicators of predictive validity and classification accuracy (Lavigne et al., 2016). Recent research has also argued against its use with preschool-age Aboriginal and Torres Strait Islander children, due to findings that did not replicate the purported five-factor structure and poor internal consistency signalling the invalidity of interpreting individual subscales (Chau et al., 2023).

The PSC is a family of measures that was originally conceived as a 35-item screening measure. It has since been adapted to include the shorter, 17-item PSC-17 (Gardner et al., 1999), Baby PSC for infants aged 1–17 months (Sheldrick, Henson, Neger, et al., 2012), and Preschool PSC (Sheldrick, Henson, Merchant, et al., 2012) along with youth report, pictorial versions and translations into other languages (Massachusetts General Hospital, 2007–2024). By spanning the ages from birth to 17 years, the PSC measures have the benefit of covering a wide range of ages. The PSC measures are simple to administer and easy to score, making them the most commonly recommended screening tool by MH authorities in the US (Semansky et al., 2003). A key difference between the SDQ and PSC is the availability of the measure in a digital format. Whilst free to use in paper-based formats, the SDQ involves costs for digital programming or scoring, which in the context of universal screening or large population studies can present a significant financial barrier. In contrast, all versions of the PSC are freely available to use in paper-based and digital formats, which is important not only for population-based screening interventions, but also for routine use of measures by practitioners to examine treatment effectiveness, such as in measurement-based care. It is particularly important for clinical and research work with parents, who may benefit from the convenience and flexibility of online measures. Online measures also expand the scope of research by making possible research with national samples of significant scale, increasing accessibility to those located in regional or rural areas (Meirte et al., 2020). Moreover, electronic methods of data collection result in less missing data and higher data quality than paper-based methods (Meirte et al.,

2020). Investigating the validity of the PSC measures in young children, in community samples, and establishing normative data may benefit researchers and clinicians looking for alternate measures of child MH.

The PSC-17 measure (see page 21) is validated for children aged 4–17 years, although it is used with children as young as two years (T. Benheim, M. Jellinek and J. Murphy, personal communication, March 7, 2023; Wagner et al., 2015). Whilst there is a strong body of evidence validating the PSC measure in older children, it has not been validated in children under four years. There is a clear need to validate the PSC-17 with parents of children aged three years, since the measure is currently in use in both clinical and research practice with this age. The use of the PSC-17 with preschool educators has been researched, even though it has not been validated for use with parents of that age range, and there is research with 4–5-year-old children in preschool settings (DiStefano et al., 2017; Gao et al., 2022; Liu et al., 2020; Moore, Dowdy, Fleury, et al., 2022). Two studies incorporated parent- and teacher-reports, whilst another two studies only utilised teacher-reports, highlighting a need for further validation of the PSC in preschool ages with parent report. Whilst evidence for the PSC-17 is strong, there remain significant gaps in the literature, particularly for young children and research outside the US.

The Preschool PSC measure (see page 22) is a version of the PSC-17 adapted with developmentally appropriate wording for young children aged 1.5–5.5 years (18–65 months). It consists of 18 items with the same three subscales as the PSC-17, in addition to a fourth subscale called "Parenting Challenges," which targets parenting difficulties across six items. Due to the low number of items per subscale, the scale developers caution against using subscale scores and instead suggest the use of the total score only (Sheldrick, Henson, Merchant, et al., 2012). This is a limiting factor of the scale for users who wish to target particular domains such as internalising problems. Although there is a growing number of studies validating the Preschool PSC, it has not been investigated to the same extent as the

original PSC-17, thus warranting further attention. Moreover, the two PSC measures currently overlap for children aged 4–5.5 years. For researchers and clinicians working with children in this specific age range, there is a lack of clarity regarding which measure is most appropriate for use.

For both PSC measures, there are gaps in the research literature, particularly in relation to multi-informant ratings. There is no research validating the Preschool PSC with educators. Validating an educator-report version of the Preschool PSC would also enable the investigation of incremental validity amongst parents and educators. Currently, only one study has investigated the incremental validity of PSC-17 ratings in relation to an outcome of kindergarten socio-emotional readiness; researchers reported a significant improvement in identification of child MH problems when educator ratings were added to parent ratings for the PSC-17 (Moore, Dowdy, Hinton, et al., 2022). Thus, incremental validity for both PSC measures is an important property to investigate in order to examine the value of multi-informant screening. Moreover, the acceptability of the PSC-17 to educators and the acceptability of the Preschool PSC to either parents or educators has not been explored.

Whilst validation evidence for both PSC measures includes diverse samples, studies have mostly been conducted in clinical populations. For both the Preschool PSC and PSC-17, the majority of studies investigating their validity have been conducted in clinical samples with primary care or hospital patients (e.g., Moreira et al., 2019; Steinbaum et al., 2008; Suman et al., 2022; Wagner et al., 2015). Currently, there is no normative data for the Preschool PSC anywhere throughout the world, and the only normative data produced for the PSC-17 is from the US (Gardner et al., 1999; Murphy et al., 2016). The establishment of normative data for measures is important for the clinical utility of screening measures in terms of enabling a comparison of individual scores to others in the sample. This highlights a need to investigate the use of the PSC measures in community populations, given a key tenet of psychological

research is generalisability to larger populations across contexts, which is especially important for population-based universal screening programs. The purpose of this research is to improve the measurement of MH in preschool-age children. By establishing validated measures of child MH that can be utilised in clinical and community child populations across Australia, this series of studies may increase the likelihood that young children at risk of MH problems will be identified, and may access interventions early in life and early in the developmental trajectory of their difficulties. If the measures are valid and acceptable by both parents and educators, there is potential for widespread use of these measures, increased early intervention and impact on the prevalence and prognosis of MH problems.

Pediatric Symptom Checklist-17 (PSC-17)
(Gardner et al., 1999)

Please mark under the heading that best fits your child.

Response options: *Never* (0), *Sometimes* (1), *Often* (2).

Scoring instructions:
1. Assign a "0" for each "never" response, a "1" for each "sometimes" response, and a "2" for each "often" response.
2. The internalising score is a sum of all internalising subscale items; the externalising score is a sum of all externalising subscale items; the attention score is a sum of all attention subscale items; and the total score is a sum of all items.

		Subscale
1	Fidgety, unable to sit still	A
2	Feels sad, unhappy	I
3	Daydreams too much	A
4	Refuses to share	E
5	Does not understand other people's feelings	E
6	Feels hopeless	I
7	Has trouble concentrating	A
8	Fights with other children	E
9	Is down on him or herself	I
10	Blames others for his or her troubles	E
11	Seems to be having less fun	I
12	Does not listen to rules	E
13	Acts as if driven by a motor	A
14	Teases others	E
15	Worries a lot	I
16	Takes things that do not belong to him or her	E
17	Distracted easily	A

Subscales: I = Internalising, E = Externalising, A = Attention

Preschool Pediatric Symptom Checklist
(Sheldrick, Henson, Merchant, et al., 2012)

These questions are about your child's behaviour. Think about what you would expect of other children the same age, and tell us how much each statement applies to your child.

Response options: *Not at all* (0), *Somewhat* (1), *Very much* (2).

Scoring instructions:
1. Determine the total score by assigning a “0” for each “not at all” response, a “1” for each “somewhat” response, and a “2” for each “very much” response, and sum the results.
2. For items where parents have selected multiple responses for a single question, choose the more concerning answer (i.e. "somewhat" or "very much").
3. A missing item counts as zero.

			Subscale
1	Does your child…	Seem nervous or afraid?	I
2		Seem sad or unhappy?	I
3		Get upset if things are not done in a certain way?	I
4		Have a hard time with change?	I
5		Have trouble playing with other children?	I
6		Break things on purpose?	E
7		Fight with other children?	E
8		Have trouble paying attention?	A
9		Have a hard time calming down?	PC
10		Have trouble staying with one activity?	A
11	Is your child…	Aggressive?	E
12		Fidgety or unable to sit still?	A
13		Angry?	E/I
14	Is it hard to…	Take your child out in public?	PC
15		Comfort your child?	PC
16		Know what your child needs?	PC
17		Keep your child on a schedule or routine?	PC
18		Get your child to obey you?	PC

Subscales: I = Internalising, E = Externalising, A = Attention, PC = Parenting Challenges

1.7 Overview and Aims of the Thesis

The overall aim of the thesis is to examine the validity of the PSC screening measures suitable for use in the preschool population, using multi-informant reports from primary caregivers such as parents and educators. A series of five studies comprehensively address this aim. First, to synthesise knowledge of existing screening measures for preschool children and their psychometric properties, Chapter 2 presents the first systematic review of multi-informant universal screening measures for preschool children, following preferred reporting items for systematic reviews and meta-analyses (PRISMA) methodology. The review answers several research questions related to existing MH measures, including what screening measures are validated for use by both parents and educators of preschool-age children. The review, importantly, evaluates the predictive and incremental validity, effectiveness and acceptability of these measures amongst parents and educators.

Having identified shortcomings in the measures available for preschool-age children, Chapters 3–5 focus on examining the psychometric accuracy of the Preschool PSC and PSC-17 for young children. Chapter 3 presents a cross-sectional study to examine the psychometric properties and normative data for the PSC-17. In this study, a large, national sample of parents of preschool- and school-age children were recruited to test the validity of the PSC-17 and to establish normative data for an Australian population. Chapter 3 addresses questions of how accurate and reliable the parent-reported PSC-17 is and whether there is a significant difference in acceptability of this measure between parents of preschool-age children (3–5 years) compared to parents of school-age children (6–17 years). Of note, this chapter examines the evidence for the validity of the PSC-17 with parents of children aged three years and presents Australian normative data for the PSC-17 for the first time. This study uses the most common criterion measure in classification accuracy studies, the CBCL, to examine concurrent and predictive validity.

Chapter 4 presents two studies. The first study complements the previous chapter by investigating the validity, reliability and acceptability of the parent-reported Preschool PSC in children aged 3–5 years. It also presents normative data for the Preschool PSC for the first time. The second study in Chapter 4 examines the psychometric properties of educator-report versions of the Preschool PSC and PSC-17, incremental value of parent and educator ratings on both PSC measures and acceptability of the PSC measures by educators—the first to do so. The study uses a paired parent and educator sample to conduct these analyses.

Having tested the validity of both PSC measures compared to comprehensive questionnaire criterion measures in earlier chapters, Chapter 5 extends the investigation of these measures by examining three important concepts, using the paired, multi-informant sample presented in the previous chapter. Through this single cohort, repeated measures study, Chapter 5 first examines the stability of child MH symptoms in a preschool community sample using the PSC screening measures. Second, the study presents the most comprehensive investigation of the concurrent and predictivity validity of the Preschool PSC and PSC-17 to date through comparison to gold-standard, diagnostic, clinical interviews and the CBCL. Thus, the study answers the question of whether parent- and educator-rated PSC screening measures correspond with clinical assessments. Finally, this study examines whether the Preschool PSC and PSC-17 differ in their level of association with clinician diagnosis and CBCL at different ages to determine whether one of the PSC measures should be preferenced over the other at a specific child age. Chapter 6 brings together the findings of earlier chapters to discuss the strengths and limitations of the research, as well as the implications of the findings for use of the PSC measures in clinical and research practice. It also discusses key directions for future research on universal screening for MH problems in the preschool period.

A Note to the Reader

Chapters 2–5 consist of published journal articles or manuscripts which have been submitted for publication and/or are currently under review. They are included in this thesis with minor amendments such as additional section headings and numbering; modified table placement and formatting to improve readability; and consistent use of language and acronyms across chapters. Articles have been included in the Appendices in their published format. Composed and published as discrete articles but brought together here as part of an overall thesis, some parts of these chapters, therefore, overlap in content. The occasional repetition in content hopefully does not diminish the reader's experience but instead signals important themes and continuation of ideas throughout the thesis as a whole.

CHAPTER 2: Multi-informant Universal Mental Health Screening for Preschool-aged Children by Parents and Educators: A PRISMA Systematic Review

The contents of this chapter are published as:

McLean, R. K., Tully, L. A., Brinley, S. K., Carl, T., Turnell, A., Northam, J. C., & Dadds, M. R. (2024). Multi-Informant Universal Mental Health Screening for Preschool-Aged Children by Parents and Educators: A PRISMA Systematic Review. *Clinical Child and Family Psychology Review*, *28*, 1–21. https://doi.org/10.1007/s10567-024-00506-2

2.1 Chapter Synopsis

Children develop rapidly in the preschool period, making accurate appraisals of MH difficult. The preschool years are a key period for early identification of MH concerns and could benefit from multi-informant, universal MH screening (UMHS). This systematic review aimed to identify multi-informant UMHS measures for preschool-age children, and to examine their clinical utility, effectiveness, and acceptability. Studies reporting the predictive and incremental validity, effectiveness or acceptability of parent- and educator-report UMHS measures for children aged 3–5 years were identified through CINAHL, Embase, ERIC, Medline, PsycINFO, Scopus and Web of Science. Studies were excluded if screening was not the primary focus, not universal, single informant, or primarily focussed on autism spectrum disorder. A total of 11 studies using 10 measures was identified. Ten studies screened for broad MH difficulties. Three educator-report and one parent- and educator-report measures had acceptable predictive validity. One study reporting incremental validity found that adding educator-report to parent ratings significantly improved the identification of MH concerns. No studies reported on effectiveness. Three studies that explored acceptability reported strong support for either UMHS in general or for specific measures. There are promising results that

UMHS can accurately identify child MH concerns in the preschool period using parent and educator reports. However, with few of the examined measures reaching the benchmark standards for predictive validity and only one study examining incremental validity, further research is needed to establish clinical utility. UMHS with preschool populations appears to be acceptable; future studies should further examine multi-informant screening in preschool populations. This systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO; https://www.crd.york.ac.uk/prospero; registration number: CRD 42022383426).

2.2 Introduction

Children develop rapidly in the first five years of life and it is well established that many MH difficulties first emerge during the preschool period from 3 to 5 years of age (Bayer et al., 2009; Lavigne et al., 2009; Oh et al., 2015; Robinson et al., 2008). Rapidly changing behaviour, social, and emotional development during this period of childhood can make accurate appraisals of child wellbeing difficult for parents (Konold et al., 2004). Parents and other caregivers of preschool-age children, such as educators or teachers (hereafter referred to as educators), face challenges in distinguishing between normative behaviour and behaviour indicative of more serious concerns (Wakschlag et al., 2007). For example, behaviours such as tantrums, non-compliance and aggression can be age-appropriate and transitory, but may also be indicative of serious, disruptive behaviour problems, especially when these behaviours are severe and/or persist over time. Thus, primary caregivers such as parents and educators need tools to help them identify children who are at risk of developing MH difficulties, to increase access to early interventions. UMHS is a population-based approach in which individuals complete systematic assessments designed to identify those at risk of developing socio-emotional (internalising) and behavioural (externalising) difficulties (Humphrey & Wigelsworth, 2016).

This review aims to examine the current literature that has examined UMHS measures in preschool children through multi-informant reports by parents and educators.

2.2.1 Early Intervention and Identification in Children

Early intervention refers to a broad array of academic, medical or social support or treatment programs designed to enhance a young child's development (Ramey & Ramey, 1998). Programs can be delivered to at-risk children, with the aim of being preventative, or targeted treatment programs implemented with children with known or diagnosed problems. Identifying children who may be at risk of MH difficulties, or recognising specific behaviour or emotional difficulties as requiring further investigation, is the first step in early intervention. Early identification of children with MH concerns often leads to intervention, thus disrupting the often stable trajectory of MH difficulties from childhood to adulthood.

Parents and other primary caregivers play a significant role in help seeking for their children, that is, the recognition of emerging MH problems and facilitating access to services (Schnyder et al., 2020). The parental help-seeking process has been identified as (1) parents' recognition of MH difficulties, (2) parents' recognition of a need for professional help, (3) parents actively seeking help, and (4) family accessing and receiving necessary help/support (Reardon et al., 2017). However, parents do not always recognise behavioural or emotional difficulties as requiring further investigation, nor do they always appreciate the need for early intervention despite the established evidence base (Kowalenko, 2012). This is particularly pertinent in the preschool-age range in which differentiating between psychopathology and normal early childhood development is a major challenge (Sim, Thompson, Marryat, Ramparsad, et al., 2019). A nationally representative survey of Australian parents of children under 18 years found that only 35% of parents were confident that they would recognise the signs of MH difficulties in their children, and fewer than half (44%) knew where to access

professional help if their child was experiencing difficulties, with confidence rates significantly lower in parents of younger than older children (Rhodes, 2017). A longitudinal study has shown that parents who recognise their child's problem behaviour are more likely to access professional help for their child at age three (Oh et al., 2015). Thus, this evidence demonstrates the importance of recognising MH concerns as the first step towards parental help seeking, including further assessment and accessing early intervention services.

Children with untreated MH difficulties can experience far-reaching negative effects throughout adolescence and into adulthood. In children as young as three, behavioural and emotional difficulties have influenced wellbeing in later childhood and early adulthood across domains such as employment, education, criminal activity, physical and MH (Caspi et al., 1996; Jones et al., 2015). However, left untreated, the implications of low rates of identification mean that children in need of MH services frequently do not receive appropriate intervention and face poor outcomes. Numerous papers have outlined the poor rates of help seeking particularly amongst parents of young children (Ellingson et al., 2004), low rates of professional service utilisation amongst preschool-age children with clinical diagnoses (Lavigne et al., 2009) or unmet MH needs in children more generally (Hiscock et al., 2020; Kataoka et al., 2002). In fact, the highest rates of unmet MH needs in the US are in children under six years, who are also children of colour and from low-income families (Kataoka et al., 2002).

2.2.2 Universal Mental Health Screening

Given overburdened healthcare systems and the long-term ramifications of untreated MH difficulties, UMHS offers a cost-effective mechanism for identifying young children in need of support (Humphrey & Wigelsworth, 2016). UMHS can be an effective tool for identifying MH concerns by improving recognition and increasing access to early intervention

services, thereby facilitating help seeking by parents or other informants (Humphrey & Wigelsworth, 2016). UMHS provides early identification of individuals with MH symptoms, who might require further follow up and it may also include referral pathways to appropriate services (Lavigne et al., 2016). It can be conducted as a single assessment with questionnaires which are scored according to clinical cut-offs, or involve multistage assessments in which children are identified in a first stage and receive further follow-up or assessment in subsequent stages. As with other standardised assessments, it can be completed individually by children, parents, educators, or other health professionals. UMHS does not make diagnoses of specific disorders and is not a diagnostic tool.

UMHS in the preschool years is important because it provides a critical opportunity to identify children at risk of MH difficulties and provide early intervention prior to difficulties becoming more entrenched and severe (Moore, Dowdy, Fleury, et al., 2022). This is especially important since early intervention, implemented in early childhood, has demonstrated high rates of cost effectiveness (Heckman, 2008). Despite the potential benefits of UMHS for preschool children, there is a paucity of research examining the efficacy, acceptability, and accuracy of UMHS in this developmental period (Anderson et al., 2019). A recent review by Becker-Haimes et al. (2020) of free, brief, and accessible child MH measures identified only two UMHS measures appropriate for use with preschool-age children which were rated as having "excellent" psychometric accuracy, the PSC and the SDQ. Of these measures, only the SDQ was also identified as suitable for multi-informant use by parents and teachers. The current review builds on the work of Becker-Haimes et al. (2020) by specifically evaluating the effectiveness and acceptability of multi-informant screening measures in the preschool-age range.

2.2.3 Multi-informant Screening

Multi-informant report is a hallmark of developmental research and practice, as it captures information about the child from multiple sources. Therefore, it is likely that UMHS for preschool-age children will be particularly effective when multiple reporters are included (Scott et al., 2011). Such approaches incorporate the unique perspectives of the child's behaviour and wellbeing across settings such as the home and school or childcare by different caregivers (Anderson et al., 2019). Given the challenges of identifying normative versus concerning emotions and behaviour in the preschool-age bracket, secondary informants such as education or childcare providers are ideal, as they may offer important and distinct insight into child MH in addition to parent perceptions (Feeney-Kettler et al., 2011; Smith, 2007). Educators are often familiar with children's social interactions, behaviour and wellbeing as they spend significant amounts of time observing children in naturalistic environments and in play with similar-aged peers. In Australia and the US, the majority of 4–5 year old children are enrolled in preschool or ECEC, attending between 15–21 hours per week on average (Australian Bureau of Statistics, 2021b; National Center for Education Statistics, 2016, 2018), again emphasising that educators are well-placed to conduct UMHS with an accessible cohort of children (DeLucia et al., 2022). However, there remains a need for children with MH difficulties to be systematically identified and referred by educators as part of an integrated model for child MH assessment (Casale & Reyes, 2023).

In contrast to primary care settings or formal MH services, preschools, pre-kindergarten, childcare, long day care or early learning centres (henceforth referred to as 'preschools') are an ideal setting for UMHS, since they are less encumbered by issues of access and stigma (Desta et al., 2017). Like schools for older children, preschools offer a unique opportunity to address the MH needs of children (von der Embse & De Los Reyes, 2024). Easy access to preschools means that many parents turn to educators or preschool staff in the first

instance to discuss MH concerns they have about their child(ren). In fact, research shows that parents frequently seek help from an informal source such as educators, friends or family, prior to seeking help from professional medical specialists, education or MH services (Pavuluri et al., 1996). Yet preschool educators have been subject to minimal research regarding their perspectives on child MH, and their role in identifying and supporting young children with socio-emotional or behavioural concerns (Croft et al., 2015). Instead, research to date on educator involvement in UMHS has focussed on primary and secondary school teachers. Research with older children has shown that over-reliance on single sources of information and decision-making based on non-scientific factors (i.e., factors which are not evidence-based, such as small standardisation samples which are not representative of the target population) contributes to inequitable school MH decision-making (von der Embse & De Los Reyes, 2024). These findings are likely to be relevant for younger children as well, therefore emphasising the need for validated, multi-informant assessment.

Identifying emotional and behavioural concerns in children can be undertaken with varying degrees of accuracy by parents and other primary caregivers, each with their own level of biases, knowledge and understanding of the child. Whilst preschool educators have a referent classroom group by which to compare the child, they may have little to no experience in the area of child MH (Stormont & Stebbins, 2005). Screening practices that rely exclusively on either parent-only or educator-only reports omit important information about child behaviour in different settings and are at risk of misidentifying children with MH concerns (Eklund & Dowdy, 2014; Ştefan & Miclea, 2017). Thus, in order to identify children who may require support and additional MH services, parents and educators need clinically useful and effective screening measures that allow for multi-informant reports to help them accurately identify children at risk.

2.2.4 Clinical Utility and Effectiveness of Screening

A range of screening tools currently exist for identifying MH difficulties in young children; however, little is known about the multi-informant measures that exist, and the clinical utility and effectiveness of these measures. The clinical utility of these measures needs to be assessed to make value judgements of UMHS in this population group. However, "clinical utility" is a poorly defined construct in the UMHS literature and has been used to encompass an array of broad concepts such as feasibility, practicability, acceptability, perceived utility and treatment effectiveness (Murphy et al., 1996; Proctor et al., 2011; Schubiner et al., 1994). This review defines "clinical utility" as whether the intended screening outcomes have perceived clinical value, that is, whether they are helpful and accurate in identifying children at risk of MH difficulties (Gall et al., 2000; Humphrey & Wigelsworth, 2016; Schubiner et al., 1994). Psychometric properties of screening measures can thus be used to assess the accuracy or validity of measures (Humphrey & Wigelsworth, 2016). When screening tools are inaccurate and thus lack clinical utility, they can incorrectly categorise children as at-risk when in fact they are not, an error that is known as false positives. Over-identification and risks of high false-positives in UMHS can result in wasted resources where children are unnecessarily triaged to receive intervention and/or stigmatisation of children who do not require help (Sawyer et al., 2014). Conversely, under-identification in the form of failing to identify children who are in fact at-risk, the error of false negatives, carries the risk of denying children access to treatment that may assist them. Over- and under-identification can be measured by the psychometric properties of predictive validity, including specificity and sensitivity. Reviewing measurement properties can be a time-consuming task requiring an understanding of psychometrics. This review's extraction of psychometric properties will, therefore, provide an accessible summary of the current literature.

In evaluating the clinical utility of multi-informant UMHS, the incremental validity of parent and educator report is of particular interest in considering the unique perspectives of each informant and what they add to an understanding of child MH. Adding educators' perspectives to parent report has the potential to increase the validity of early identification, whilst also raising the possibility of discrepant ratings between informants (Croft et al., 2015; De Los Reyes et al., 2015). The presence of multi-informant discrepancy has been well-reported and in fact offers important, domain-relevant information specific to each informant's context e.g., home or school (De Los Reyes et al., 2023). However, the incremental, predictive validity of multi-informant approaches relative to the use of single informant report warrants further attention, especially in relation to young, preschool-age children.

Evaluating the effectiveness of an intervention can be assessed by its ability to achieve its intended outcomes (Andrews, 1999). More specifically, in the context of UMHS, effectiveness can be defined by the improvement of early identification of child MH concerns (Brinley et al., 2024). UMHS is thus considered effective if it improves the identification or diagnosis, referral, or treatment of child MH difficulties, which can be measured by rates of accurate identification of MH difficulties or risk, uptake of referral rates to MH services, or service uptake.

Whilst there are a number of previous reviews of screening by educators for Autism Spectrum Disorder in children, including preschool-aged children (e.g., DeLucia et al., 2022), a recent systematic review examining the effectiveness and cost-effectiveness of school-based identification models including UMHS has found there is little research on the identification of MH difficulties in preschool-age children (Anderson et al., 2019). As such, there is a need to examine the literature regarding UMHS effectiveness in terms of service referral and uptake following identification.

2.2.5 Acceptability of Screening

The acceptability of UMHS can impact uptake and implementation of screening in clinical and community settings (Harrison et al., 2013). It has been argued that UMHS in young children is inappropriate and ineffective, may increase parental anxiety, stigmatise children with labels, and lead to the over-medication of children (Frances, 2012; Jureidini & Raven, 2012). These arguments made against UMHS have in some instances halted the implementation of UMHS in preschool-age children previously (Alexander & Mazza, 2015). As such it is essential to investigate the perceived acceptability of UMHS amongst users if it is to be adopted and implemented widely as an early intervention strategy.

The current review is primarily interested in users' perspectives of acceptability, that is, those of parents and educators. One element of acceptability comprises whether screening is considered appropriate, and/or perceived as helpful or useful. This may involve perceptions about whether the screening is wanted, needed or socially significant (Humphrey & Wigelsworth, 2016) or whether it is agreeable, palatable, or satisfactory (Proctor et al., 2011). Satisfaction can also encompass 'usability' of the tool (e.g., level of satisfaction with the length of screening, whether the language was easy to understand); whether informants felt comfortable or distressed by the screening; and whether informants would recommend screening to others, or if they would complete screening again in future (Brinley et al., 2024). Acceptability data, expressed as parental or educator attitudes towards screening, can be collected quantitatively or qualitatively. Importantly, acceptability encompasses characteristics that are associated with the likelihood of adoption of a screening measure (Albers et al., 2007; Kamphaus et al., 2007). This review will investigate the acceptability of UMHS broadly, and specific screening measures, as assessed by parents and educators.

No systematic review to date has evaluated multi-informant UMHS measures by parents and educators of preschool-age children. As such, there is a need to identify what measures

exist for children in this age range, examine the clinical utility and effectiveness of these measures, and also to examine the acceptability of UMHS more broadly.

2.2.6 The Current Study

The aim of this systematic review is to examine multi-informant UMHS for preschool-age children (3–5 years) by both parents and educators. The review will answer the following questions:

1. What are the existing MH screening measures utilised by both educators and parents of preschool-age children and what is the clinical utility of these measures; that is, what is the predictive and incremental validity of these measures?
2. What is the effectiveness of UMHS?
3. What is the acceptability of utilising UMHS amongst educators and parents of preschool-age children?

2.3 Method

The protocol for this systematic review was registered with the international Prospective Register of Systematic Reviews, PROSPERO, on 22 December, 2022 (Registration number CRD42022383426).

2.3.1 Eligibility Criteria

Peer-reviewed studies examining UMHS for child MH or socio-emotional and behavioural wellbeing in preschool children aged 3 years 0 months—5 years 11 months were included. Included studies reported universal screening whereby all children within a

population-based sample undertook the systematic assessment of MH or socio-emotional or behavioural wellbeing and ratings were provided by both parents and educators.

Studies were excluded if they focussed specifically on Autism Spectrum Disorder or were not published in English. Complete inclusion and exclusion criteria are presented in Table 1.

Information Sources

Seven databases were searched for relevant literature: CINAHL, Embase, ERIC, Medline, PsycINFO, Scopus and Web of Science. All databases were searched via Ovid aside from CINAHL, Scopus and Web of Science.

Table 1.

Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria (in ranked order)
	Duplicate studies
Studies published in English	Studies not published in English
Studies published in peer-reviewed journals	Studies not published in peer-reviewed journals; grey or unpublished literature (dissertations, theses)
Full text available	No full text available
Empirical data presented	No empirical data presented (Systematic review/meta-analysis; Case study; Narrative review; Conference proceedings; Commentary; Protocol paper)
Studies focussed on individuals in the age range 3–5 years (mean age ≤ 3 years 0 months and ≥ 5 years 11 months, or more than half the sample within the age range)	Studies focussed on individuals outside of the age range 3–5 years (mean age ≤ 3 years 0 months and ≥ 5 years 11 months, or more than half the sample outside the age range). This includes excluding any studies in which:
	a. No age details for child samples are reported and individuals are not attending preschool or similar early childhood learning or childcare;
	b. No age details for child samples are reported and children are enrolled in kindergarten, elementary grade or above (e.g., grade one; secondary school etc);
	c. Sample (or sub-sample analyses) do not report mean and age range includes children under 2 years and 0 months or children over 6 years and 11 months
	Screening with wrong primary respondent such as Medical Professional (General Practitioner, Paediatrician, Nurse, etc), Allied Health (Occupational Therapist, Social Work, etc), Psychologist, or Child report
Screening with parent and educator report	Screening with parent report only or educator report only
Studies in which the primary focus is screening	Studies in which the primary focus is not screening (e.g. prevalence surveys)
Studies in which the screening is universally implemented	Studies in which the screening is not universally implemented (i.e., screening is implemented with targeted populations)
Screening with a primary focus on child MH or socio-emotional and behavioural wellbeing	Screening with a primary focus other than child MH or socio-emotional and behavioural wellbeing (e.g., parenting, parental MH, physical health, academic performance)

Inclusion criteria	Exclusion criteria (in ranked order)
	Studies which do not report predictive validity, incremental validity, effectiveness or acceptability data for screening measures
	Studies which focus on screening for Autism Spectrum Disorder

2.3.2 Search Strategy

The first author developed and translated the search strategy according to the relevant subject headings and functionality of each database in consultation with two authors (LT and MD) and a research librarian. The strategy was trialled in October and December 2022. Databases were searched in May 2023 using the following search strategy:

1. (exp mass screening/ or mass screening*.mp. or universal screening*.mp. or early intervention*.mp. or early identification*.mp. or screening*.mp. or exp needs assessment/ or needs assessment*.mp OR (mass adj3 screening*) OR (universal adj3 screening*)).
   AND
2. (exp mental health/ OR mental health.mp OR exp mental disorder/ OR wellbeing.mp OR wellbeing.mp OR well-being.mp OR ((mental or psychological or behavioral or socio-emotional or social?emotional or socioemotional) adj2 (disorder* or difficult* or health* or problem* or wellbeing or well-being)).mp).
   AND
3. (preschool student*.mp OR exp Child, Preschool/ OR preschool*.mp OR pre-kindergarten*.mp OR prekindergarten* OR early childhood*.mp OR early learning*.mp). AND
4. (exp school teachers/ OR teacher*.mp OR educator*. mp OR child care worker*.mp OR day care worker*. mp OR daycare worker*.mp OR childcare worker*.mp OR "prekindergarten teacher*" OR "pre-kindergarten teacher*").

Results were limited to peer reviewed journals.

To ensure literature saturation, we scanned the reference lists of included studies identified through the search and conducted manual handsearching.

2.3.3 Study Selection

Article citations from the search results were uploaded to Covidence (Veritas Health Innovation). After duplicates were removed, the first author screened all papers by reviewing the title and abstract of papers and a team of five authors (SB, TC, RM, LT and AT) independently screened papers against criteria. A calibration exercise was undertaken to pilot and refine the screening questions. Studies not meeting criteria were excluded.

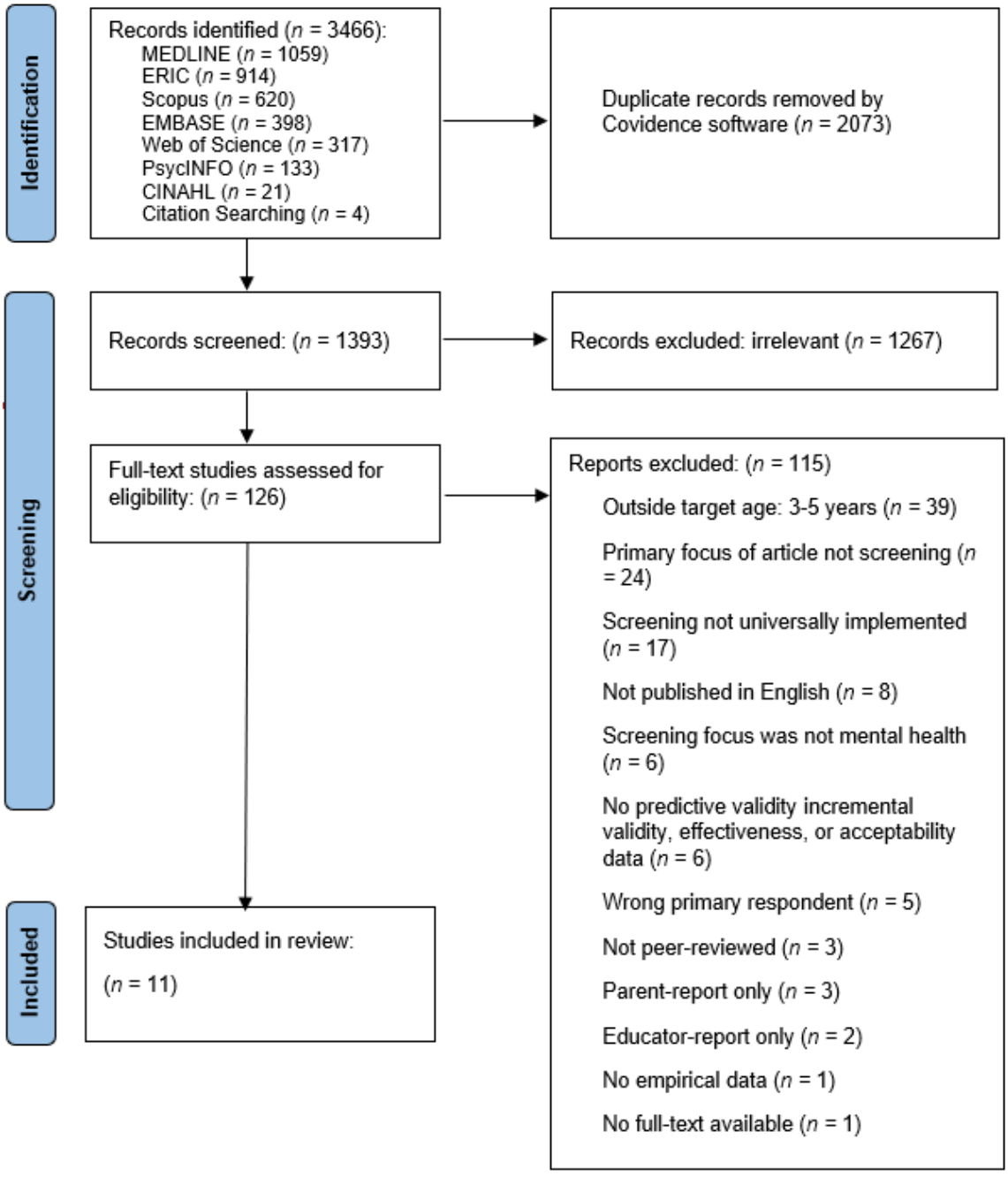
Full texts for the remaining studies were obtained and examined against criteria. Studies not meeting the criteria were excluded using a hierarchy of exclusion criteria as outlined in Table 1. All papers were screened by two reviewers to reduce selection bias. The first author screened all papers and the same team of five authors independently screened a selection of papers against criteria. At either stage of screening, a third reviewer was consulted when consensus could not be reached by prior discussion. The study selection process is outlined in Figure 1.

Data Collection

Data for the included studies were extracted in Excel using a codebook developed by the authors. Six authors (SB, TC, RM, JN, LT and AT) independently conducted data extraction. Extraction into the codebook was trialled twice using two included studies. Codebook headings are included in Appendix A-1. Our data collection only utilised published data and we did not contact study authors for additional information.

Figure 1.

PRISMA Flow Diagram for Identification of Studies



2.3.4 Study Quality Assessment

All included studies were critically appraised for risk of bias using the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018), an evaluation tool designed for systematic reviews that include mixed studies i.e., qualitative, quantitative or mixed methods studies. The MMAT comprises two stages. First, two initial screening questions confirm that the study is an empirical study and appropriate to be appraised by the MMAT. Second, for each included study, the appropriate methodological category (e.g., quantitative descriptive, quantitative non-randomised, qualitative) is selected before answering corresponding questions. A calibration exercise was undertaken to ensure all reviewers understood how to assess studies using the MMAT and resulted in an acceptable level of agreement between raters.

2.4 Results

Database searches yielded 3,466 results. After duplicates were removed, 1,393 were screened including 126 that underwent full-text review. The final number of studies included was 11. The study selection process is illustrated in Figure 1.

2.4.1 Study Characteristics

All studies were published within the past 10 years, aside from two studies published in 2007 (Barbarin) and 2011 (Feeney-Kettler et al.). Of the 11 studies included in the review, over half of the studies were conducted in the US (n = 7). Studies were also conducted in Japan (n = 1), the Netherlands (n = 1), Romania (n = 1), and Spain (n = 1). There were no studies conducted in low and middle income countries.

Mental health domains screened in the included studies were mainly broad social, emotional, and behavioural wellbeing (n = 10), and ADHD (n = 1). All children were

community samples recruited from either preschools or kindergartens (n = 8), community health clinics (n = 2) or random census sampling (n = 1). Of the studies reporting demographic details about parents, 70% of recruited samples were predominantly female. Of the studies that reported child gender (n = 6), all recruited samples were gender balanced i.e., between 45–55% female/male. No study reported non-binary gender for children. The socio-economic status and cultural diversity of samples varied across studies. A summary of study characteristics is provided in Table 2.

Table 2.

Study Characteristics

Author (Year)	Country	Study Design	Population & Setting	Study Aims	Child Characteristics		
					Sample size (*n*)	Gender (female, male)*	Age range (*M*, *SD*)
Barbarin et al., (2007)	USA	Cross-sectional analytical study	Study 1: Pre-K children in 40 randomly selected classrooms in 6 participating states Study 2: Children enrolled in the Head Start program or early childhood programs receiving partial state financial support	To determine the nature and prevalence of socio-emotional concerns parents and educators have about preschool children, the degree to which parents and educators agree on their concerns, and the psychometric properties of the ABLE	Study 1: 415 Study 2: 5,577	Study 1: NR Study 2: NR	Study 1: 4 yrs (NR) Study 2: 3–4 yrs (NR)
Doove et al., (2019)	Netherlands	Prospective observational study	Children aged 3 years at study entrance as part of the Monitoring Outcome Measurements (MOM) child development study	To define psychometric properties of the Dutch PEDS and three VAS about 'parenting', 'child behaviour' and 'child competence' at the age of 3 and 4 years	346	52%, 48%	3 years (3.0; 0.2)
Ezpeleta et al., (2013)	Spain	Cross-sectional analytical study	A random sample of 3-year-olds from the census of preschoolers in Barcelona	To evaluate the psychometric properties of the SDQ in sample of Spanish preschool children	Phase 1: 1,341 Phase 2: 622	Phase 1: NR, 50.9% Phase 2: NR, 50%	Phase 1: 3 years (3.0; 0.18) Phase 2: 3 years

Author (Year)	Country	Study Design	Population & Setting	Study Aims	Child Characteristics		
					Sample size (*n*)	Gender (female, male)*	Age range (*M*, *SD*)
							(3.0, 0.16)
Feeney-Kettler et al., (2011)	USA	Quantitative Descriptive	Children in *n* = 22 preschools and childcare facilities in Southern California and Tennessee	To streamline the PBSS Phase 2 to create separate parent & educator versions, and to evaluate psychometric properties of the PBSS as a cost-efficient universal screening tool	NA	49%, 51%	3–5 years (NR)
Feeney-Kettler et al., (2019)	USA	Quantitative Descriptive	Preschool children recruited from day cares and preschools (number not reported) in a large southern city and its surrounding suburbs in the US	To evaluate the multiple-gate PBSS for identifying children's social, emotional, and behavioural difficulties	122	48%, 44%	3–5 years (NR)
Girio-Herrera et al., (2015)	USA	Cross-sectional analytical study	Kindergarteners at *n* = 18 elementary schools (Study 1) *n* = 5 elementary schools (Study 2)	To examine the IRS as a screening tool for detecting kindergarten children who are at risk for social, emotional, academic, and behavioural difficulties	NA	Study 1: NR, 46.8% Study 2: NR, 50.8%	5 years 1 month–5 years 11 months Study 1: (5.48, 0.32) Study 2: (5.61, 0.46)

Author (Year)	Country	Study Design	Population & Setting	Study Aims	Child Characteristics		
					Sample size (*n*)	Gender (female, male)*	Age range (*M*, *SD*)
Kettler et al., (2017)	USA	Cross-sectional analytical study	Preschool children taken from a convenience sample of *n* = 33 preschools in a heavily populated Northeastern US state	To determine the internal consistency, cross-informant agreement of a multi-gate screener and the concurrent relations among scores from single-gate and multi-gate screeners	105	NR, 60%	3–5 years (NR)
Moore, Dowdy, Hinton, et al. (2022)	USA	Quantitative Descriptive	Preschool, Kindergarten, and first-grade children from *n* = 5 state-funded preschools housed within public elementary schools	A preliminary examination of educators' and parents' beliefs related to their participation in Universal Mental Health Screening	NA	NR, NR	NR (NR)
Moore, Dowdy, Fleury, et al. (2022)	USA	Cross-sectional analytic and longitudinal cohort study	Preschool children from *n* = 5 state-funded preschools from a Title 1 (low income supported) school district	To examine the use of informants at the preschool level when universally screening for behavioural and emotional risk	535	48.3%, 44%	3–5 years (4.32, NR)
Ştefan et al., (2017)	Romania	Cross-sectional analytical study	Preschool children from *n* = 3 preschools within an urban area of Cluj-Napoca	To evaluate the reliability and validity of the ECS and SCS parent and educator versions	Sample 1: 107 Sample 2: 73	Sample 1: 48.6%, 51.4% Sample 2: 54.8%, 45.2%	Sample 1: 2 years, 4 months–4 years (43.84 months, NR)

Author (Year)	Country	Study Design	Population & Setting	Study Aims	Child Characteristics		
					Sample size (*n*)	Gender (female, male)*	Age range (*M*, *SD*)
							Sample 2: 4–5 years (4.33, NR)
Takayanagi et al., (2016)	Japan	Quantitative Descriptive	5-year-old children from a local community health check-up run through municipal health centres in Aomori prefecture	To verify the psychometric properties of the ADHD-RS in identifying preschool children with ADHD using DSM-5 criteria at a community health check-up for 5-year-old children	838	46.1%, 53.9%	4 years. 10 months–5 years, 10 months (4.83, 0.28)

Notes. *no studies reported non-binary child gender. ABLE = Attention Behaviour Language Emotions, ADHD = Attention-Deficit/Hyperactivity Disorder, ADHD-RS = Attention-Deficit/Hyperactivity Disorder Rating Scale, DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, ECS = Emotion Competence Screening, IRS = Impairment Rating Scale, NR = not reported, PBSS = Preschool Behavior Screening System, PEDS = Parents' Evaluation of Developmental Status, SCS = Social Competence Screening, US = United States, VAS = Visual Analogue Scales.

Results of Individual Studies

2.4.2 Overview of Measures

The research studies examined 10 measures. As is expected for universal screening measures of community populations, almost all measures assessed broad constructs of internalising and externalising difficulties, including the Attention, Behavior, Language, Emotions (ABLE; Barbarin, 2007); Behavior Assessment System for Children (BASC; Moore, Dowdy, Fleury, et al., 2022; Moore, Dowdy, Hinton, et al., 2022); Behavior Assessment System for Children–Behavioral and Emotional Screening System, (BASC-2 - BESS; Kettler et al., 2017); Emotion Competence Screening (ECS; Ştefan & Miclea, 2017); Impairment Rating Scale (IRS; Girio-Herrera et al., 2015); Parents' Evaluation of Developmental Status (PEDS; Doove et al., 2019); PSC-17 (Moore, Dowdy, Fleury, et al., 2022); Preschool Behavior Screening System (PBSS; Feeney-Kettler et al., 2019; Feeney-Kettler et al., 2011; Kettler et al., 2017); and SDQ (Ezpeleta et al., 2013). One measure focussed on attention or hyperactivity, ADHD-Rating Scale-IV (Takayanagi et al., 2016). Table 3 details the measures used in the studies. Some measures also included prosocial or adaptive behaviour (ECS, IRS, PBSS, SDQ); and the PEDS measured developmental domains in addition to social, emotional and behavioural constructs (Doove et al., 2019). The number of items in each screening measure varied widely, particularly as some measures were multiphasic and included additional items dependent on previous responses (e.g., for each issue that a respondent indicated concern, follow-up questions were then asked about the level of severity and impact on the child's life).

Most studies reported inter-rater reliability between parent and educator informants using Pearson's correlations. Two studies reported moderate cross-informant agreement between parent and educator ratings (Barbarin, 2007; Feeney-Kettler et al., 2011). Barbarin (2007) reported that parents and educators agreed 77% of the time about children who did not have difficulties. One study presented chi-square analyses of cross-informant agreement,

reporting a statistically significant association between parental concerns and professional caregivers' concerns about child wellbeing and development at baseline and 10-month follow-up ($\chi^2 = 34.8$; $df = 1$, $p < .001$ and $\chi^2 = 8.1$; $df = 1$, $p = .004$) (Doove et al., 2019).

The majority of studies (n = 7) that reported internal consistency for parent-report measures reported high reliability (Cronbach's alphas above .80) (Ezpeleta et al., 2013; Feeney-Kettler et al., 2019; Feeney-Kettler et al., 2011; Kettler et al., 2017; Moore, Dowdy, Fleury, et al., 2022; Ştefan & Miclea, 2017; Takayanagi et al., 2016). Most studies ($n = 7$) that reported internal consistency for educator-report measures reported high or very high reliability (Cronbach's alphas above .80) (Ezpeleta et al., 2013; Feeney-Kettler et al., 2019; Feeney-Kettler et al., 2011; Kettler et al., 2017; Moore, Dowdy, Fleury, et al., 2022; Ştefan & Miclea, 2017; Takayanagi et al., 2016). Three studies did not report internal consistency for either parent or educator report (Barbarin, 2007; Girio-Herrera et al., 2015; Moore, Dowdy, Hinton, et al., 2022).

Few studies reported test-retest reliability (Doove et al., 2019; Feeney-Kettler et al., 2011; Ştefan & Miclea, 2017). Of the three studies that did, all reported good to excellent reliability for parent report and three studies reported good to excellent reliability for educator report. See Appendix A-2 for further details regarding screening measures' inter-rater reliability, internal consistency, and test-retest reliability.

2.4.3 Clinical Utility

To assess clinical utility, the predictive and incremental validity of measures were extracted. Studies utilised a range of criterion measures to test the predictive validity of their screening measures. Criterion measures varied widely, with only one study utilising gold standard diagnostic interviews (Ezpeleta et al., 2013) or comprehensive assessments ($n = 6$) such as the CBCL (Achenbach, 1999) or Caregiver-Teacher Rating Form (C-TRF; Achenbach

& Rescorla, 2000), whilst others utilised other brief screening measures such as the PSC-17 or SDQ (n = 5) or a combination of measures. See Table 3 for details.

Seven out of the eight studies examining predictive validity reported acceptable or good sensitivity or specificity for parent ratings. Kettler et al. (2017) reported the strongest predictive validity for children in the clinical range: sensitivity 90%, specificity 90%, PPV 42% and NPV 99%. This study tested the PBSS against a comprehensive range of criterion measures including the BASC-2 - BESS, C-TRF and CBCL 1.5–5, which included both parent and educator report.

Kettler et al. (2017) also reported the strongest predictive validity for educator ratings for children in the clinical range: sensitivity 97%, specificity 91%, PPV 45% and NPV 100%. When compared to the criterion measure, educators were slightly more accurate raters than parents.

The sensitivity of the SDQ was poor for both parent and educator reports across all subscales aside from the prosocial subscale (Ezpeleta et al., 2013). One study reported combined predictive validity for parent and educator reports incorporating multi-informant reports to identify children at risk of internalising and externalising difficulties, with the PBSS showing combined sensitivity 98% and specificity 28% (Feeney-Kettler et al., 2011). Three studies did not report predictive validity (Feeney-Kettler et al., 2019; Moore, Dowdy, Fleury, et al., 2022; Moore, Dowdy, Hinton, et al., 2022)

The majority of studies (n = 8) reported Area Under Receiver Operating Characteristic Curve (AUC) analyses as an indicator of screening accuracy. Criterion measures again varied widely, as did the level of analyses conducted (e.g., informant report, subscale, child gender etc). All studies reporting AUC analyses for parent report (n = 7), reported fair (above .70) to excellent (.90–1.00) AUCs. Five studies reporting AUC analyses for educator report, reported very good to excellent AUCs (.80–1.00) (Doove et al., 2019; Feeney-Kettler et al., 2019;

Feeney-Kettler et al., 2011; Kettler et al., 2017; Ştefan & Miclea, 2017). Two studies reported AUCs that were below acceptable levels for educator report (Ezpeleta et al., 2013; Takayanagi et al., 2016). Finally, only one study (Girio-Herrera et al., 2015) tested cross-informant ratings by parent-reported Impairment Rating System (IRS) identifying educator-rated BASC 2 at-risk status, and educator IRS identifying parent-based BASC 2 at-risk status–both resulting in poor AUCs.

2.4.4 Incremental Validity

Only one study reported the incremental validity of parent and educator reports in terms of child social-emotional outcomes (Moore, Dowdy, Fleury, et al., 2022). Using hierarchical regression modelling, this study reported that educator-report of social-emotional difficulties using the BASC-3 BESS Behavioral and Emotional Risk Index (BERI) was significantly associated with kindergarten social-emotional readiness (β = -.46); however, including parent-report BESS BERI did not significantly improve prediction of kindergarten social-emotional readiness. However, when parent-report was entered in the first block, parent-report was significantly associated with kindergarten social-emotional readiness (β = -.19) and when educator ratings were added in the second block, educator-report was also significantly associated with kindergarten social-emotional readiness (β = -.44) although parent ratings in the second block were no longer significantly related to readiness (β = -.07, p =.26).

In addition, Moore, Dowdy, Fleury, et al. (2022) reported incremental validity using a second screening measure, the PSC-17. Educator-reported social-emotional difficulties using the PSC-17 was significantly associated with kindergarten social-emotional readiness (β = -.43) and parent report was not associated with kindergarten social-emotional readiness, nor did inclusion result in significant improvements in variance in the model (β = .05, R^2 = 0). When parent report was entered in the first block, it was not statistically significant. However, when

educator ratings were added to subsequent blocks, educator ratings were significantly associated with kindergarten social-emotional readiness (β = -.44) and parent ratings were no longer significantly related to social-emotional readiness. It is important to note that kindergarten social-emotional readiness is a criterion variable that references child behaviour in the education context only, and therefore, criterion contamination (whereby differences in parent and educator ratings may have arisen due to educators' influence over the criterion variable, in this case social-emotional readiness) may be a factor in these findings which seemingly favour educator report. As Moore, Dowdy, Fleury, et al. (2022) conclude, both parents and educators offer valuable information when screening child MH risk. However, educator report may be more informative than parent report for predicting kindergarten social-emotional readiness, which we note is an education-specific measure of MH. It is also important to note that this study did not report other psychometrics of interest, such as predictive validity more generally.

Table 3.

Screening Measure Properties

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
Barbarin et al., (2007)	Study 1: 238, 415 Study 2: NR, NR	ABLE (Stage 1: 10, Stage 2: 40)	Attention, Behaviour, Language, Emotions	Study 1 TCRS; BPI; ORCB; PPVT; OWLS Study 2: N/A	Educator Pre-K: .42 K: .65 Parents: .59 (end of Pre-K) .60 (end of K)	Parent: .86 (end of Pre-K), .84 (end of K)	NR, NR	NR, NR
Doove et al., (2019)	294, 329	VAS (3) PEDS-Dutch Version (10)	School Readiness/ Social Participation	CBCL SDQ	Educator: Child competence VAS 82.8 PEDS: 96.6 Parent: Parenting VAS: 90.9 Child behaviour VAS: 71.4	Educator Child competence VAS 82.8 PEDS: 83.1 Parent: Parenting VAS: 78.4	NR, NR	Educator Child competence VAS 97.6 PEDS: 99.5 Parent Parenting VAS: 99.2

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
					PEDS: 80.6	Child behaviour VAS 80.8 PEDS: 80.2		Child behaviour VAS 95.9 PEDS: 98.9
Ezpeleta et al., (2013)	Phase 1 N/A, 1,341 Phase 2 94, 622	SDQ (25)	Behaviour Difficulties	CBCL (1.5–5 years); CGAS; DICA-PPYC; SDQ (Spanish and Catalan versions)	Educator Screening cutoff: 72.1 Borderline: 38.2 Abnormal: 18.5 Parent Screening cutoff: 74.2 Borderline: 52.1 Abnormal: 29.9	Educator Screening cutoff: 44.3 Borderline: 78.3 Abnormal: 89.2 Parent Screening cutoff: 62.1 Borderline: 83.4 Abnormal: 95.8	NR, NR	NR, NR
Feeney-Kettler et al., (2011)	112 113	PBSS (46)	Internalising Externalising	BASC-2	Educator Tot .94 Int .96 Ext 1.00	Educator Tot .51 Int .59 Ext .71	Educator Tot .49 Int .44 Ext .54	Educator Tot .95 Int .98 Ext 1.00

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
			g Prosocial Behaviour		Combined Ed/P Tot .98 Int .89 Ext .98 Parent: Tot .96 Int .80 Ext .90	Combined Ed/P Tot .28 Int .42 Ext .48 Parent: Tot .49 Int .61 Ext .59	Combined Ed/P Tot .57 Int .62 Ext .57 Parent: Tot .39 Int .57 Ext .51	Combined Ed/P Tot .93 Int .77 Ext .97 Parent: Tot .96 Int .82 Ext .93
Feeney-Kettler et al., (2019)	122 122	PBSS (46)	Internalising, Externalising, Prosocial Behaviour	BASC-2	NR, NR	NR, NR	NR, NR	NR, NR
Girio-Herrera et al., (2015)	Study 1: 56 12 Study 2: 568 273	IRS (31)	Study 1: Academic, Social, Behaviour, Family Study 2: Academic, Social, Behaviour, Family	Study 1: BASC-2 Study 2: BESS	Study 1: Educator .57 Parent .14 Study 2: Educator .72 Parent .17	Study 1: Educator .92 Parent .98 Study 2: Educator .95 Parent .95	Study 1: Educator .65 Parent .98 Study 2: Educator .65 Parent .95	Study 1: Educator .89 Parent .74 Study 2: Educator .97 Parent .90

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
Kettler et al., (2017)	105 105	PBSS (46)	Social, Emotional, and Behaviour Difficulties	BASC-2, C-TRF, CBCL	P1+P2 Educator ASEBA TP Borderline .76, Clinical .97 Parent ASEBA TP Borderline .66, Clinical .90	P1+P2 Educator ASEBA TP Borderline .96, Clinical .91 Parent ASEBA TP Borderline .91, Clinical .90	P1/P2 Educator ASEBA TP Borderline .80, Clinical .45 Parent ASEBA TP Borderline .91, Clinical .90	P1/P2 Educator ASEBA TP Borderline .95, Clinical 1.00 Parent ASEBA TP Borderline .95, Clinical .99
Moore, Dowdy, Hinton, et al. (2022)	40 330	BASC-3 BESS (86)	Behavioural and Emotional Risk	PSC-17	NR, NR	NR, NR	NR, NR	NR, NR
Moore, Dowdy, Fleury, et al. (2022)	14 535	BASC-3 BESS (86); PSC-17 (17)	Behavioural and Emotional Risk	BASC-3 BESS	NR, NR	NR, NR	NR, NR	NR, NR
Ştefan et al., (2017)	NR 180	ECS (30)	Social and Emotional Competencies	SRSS	Educator ECS-T to C-TRF: Int 70.00–75.00 Ext 75.00–83.33	Educator ECS-T to C-TRF: Int 92.78–93.85 Ext 92.21–94.74	Educator ECS-T to C-TRF: Int 50.00–60.00 Ext 45.45–64.29	Educator ECS-T to C-TRF: Int 96.53–96.77 Ext 96.77–98.61

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
					SCS- T to C-TRF: Int 75.00–77.78 Ext 71.43–75.00 Parent ECS-P to CBCL: Int 80.00–84.62 Ext. 69.23–71.43 SCS-P to CBCL: Int. 76.92–80.00 Ext. 71.43–75.00	SCS-T to C-TRF: Int 91.84–94.20 Ext 90.79–91.58 Parent ECS-P to CBCL: Int 92.65–93.62 Ext 92.42–92.47 SCS- P to CBCL: Int 89.36–92.65 Ext 88.42–93.94	SCS-T to C-TRF: Int 42.86–46.67 Ext 41.67–52.94 Parent ECS-P to CBCL: Int 44.44–64.71; Ext 50.00–56.25 SCS-P to CBCL: Int 44.44–50.00 Ext 45.00–55.56	SCS-T to C-TRF: Int 97.83–98.48 Ext 96.67–97.18 Parent ECS-P to CBCL: Int 97.78–98.44 Ext 95.56–96.83 SCS-P to CBCL: Int 96.55–98.44 Ext 96.55–96.88
Takayanagi et al., (2016)	NR 838	ADHD-Rating Scale-IV Parent (18) ; Educator (18)	ADHD	SDQ	Educator Above 90th percentile 30.23 Parent Above 90th percentile 89.13	Educator 90.92 Parent 94.07	Educator 16.05 Parent 46.59	Educator 95.78 Parent 99.33

*Note*s. Data reported for recommended cutoff scores or best trade-off between sensitivity and specificity. ABLE = Attention Behavior Language Emotions; ADHD = Attention-Deficit/Hyperactivity Disorder; ASEBA = Achenbach System for Empirically-Based Assessment; BASC-2/3 = Behavior Assessment System for Children Second/Third edition; BESS = Behavioral and Emotional Screening System; BPI = Behavior Problem Index; CBCL = Child Behavior Checklist; CGAS = Children's Global Assessment Scale C-TRF = Caregiver-Teacher Rating Form; DICA-PPYC = Diagnostic Interview for Children and Adolescents for Parents of Preschool And Young Children; ECS-P/T= Emotion Competence Screening Parent/Teacher Form; Ext = externalising; K = Kindergarten; Int = internalising; IRS = Impairment Rating Scale; NR = not reported; ORCB = Observer ratings of Child Behavior; OWLS = Oral and Written Language Scale; PBSS = Preschool Behavior Screening System; PEDS = Parents' Evaluation of Developmental Status; PPVT = Peabody Picture Vocabulary Test; PSC-17 = Pediatric Symptom Checklist-17; SCS-P/T = Social Competence Screening Parent/Teacher Form; SDQ = Strengths and Difficulties Questionnaire; SRSS = Social Skills Rating System; Tot = total score; TCRS = Teacher Child Rating Scale; VAS = Visual Analogue Scales.

2.4.5 Effectiveness

Only one study reported the effectiveness of the screening measure in terms of referral uptake and longitudinal outcomes for children after screening preschool children using the ABLE. However, this study only focussed on parent reports, and did not include the effectiveness of educator report or multi-informant data. Moreover, this study reported effectiveness in terms of referral for Individualised Education Plans (IEP), the same outcome used to test predictive validity of the ABLE. After conducting screening for preschool children, 13.1% of children who were identified as having serious concerns through parent-reported ABLE were referred for IEP by the end of kindergarten. This was compared to 3.7% of children for whom no concerns were identified. The study did not report if this was a significant difference. The remaining studies did not report effectiveness data about screening measures.

2.4.6 Acceptability of Universal Mental Health Screening

Only three studies examined the acceptability of screening measures or UMHS in general amongst parents and educators. Two studies investigating the acceptability of the PBSS measure found that 88% of parents in one sample (n = 107) and 87% of parents in another sample (n = 91) reported that the PBSS could provide useful information about their child, and 91% (n = 111) and 94% (n = 99) found it clearly written (Feeney-Kettler et al., 2019; Kettler et al., 2017). Amongst educators, 88% (n = 107) and 71% (n = 74) indicated that the PBSS could provide useful information about preschool children, and 89% (n = 109) and 88% (n = 83) found it clearly written (Feeney-Kettler et al., 2019; Kettler et al., 2017).

The initial study reported that just over half of educators (55%, n = 58) would be likely to use the PBSS to screen all their students for social and emotional wellbeing and some educators indicated that they would select students to screen rather than implement screening

universally (Kettler et al., 2017). The second study reported that 71% of educators (n = 87) indicated that they would be likely to use the PBSS to screen all students (Feeney-Kettler et al., 2019). Neither study reported the number of items in the acceptability or evaluation measure.

A third study measured acceptability of UMHS more generally, using four items encompassing constructs of importance, usefulness, willingness and appropriateness (Moore, Dowdy, Hinton, et al., 2022). On average, educators and parents agreed or strongly agreed that it is important for schools to ask questions about child MH, screening is useful for identifying at-risk children, and MH development should be addressed in school (Moore, Dowdy, Hinton, et al., 2022). Only 0.3%–1.8% (n = 6) of parents and 0%–10.0% (n = 4) of educators disagreed or strongly disagreed with the acceptability of UMHS (Moore, Dowdy, Hinton, et al., 2022).

2.4.7 Risk of Bias in Studies

All studies were assessed using the MMAT (Hong et al., 2018). All studies were considered moderate to high quality. Quantitative descriptive studies tended to rate poorly for the risk of non-response bias or explaining non-responders and low response rates. Non-randomised trials frequently did not adequately account for confounders in the design or analysis. There were no randomised trials identified.

2.5 Discussion

The aim of this review was to systematically examine research conducted on multi-informant UMHS for preschool-age children by both parents and educators. We identified which measures have been utilised in the preschool population and examined their clinical utility in terms of predictive and incremental validity and effectiveness. The review also

examined the acceptability of these measures amongst parents and educators. The recency of included studies suggests that this is an emerging area of research, conducted mainly in the US and in high-income, industrialised nations. Identified screening measures were highly varied and for some aspects examined, such as incremental validity, effectiveness and acceptability, there was a paucity of research making definitive conclusions difficult. However, there were 11 included studies and 10 measures identified for multi-informant UMHS in the preschool period, with promising findings for some measures suggesting that UMHS can accurately identify child MH concerns using parent and educator report.

In order to assess the clinical utility of screening, this review examined the predictive and incremental validity of MH screening measures. Several screening measures demonstrated strong predictive validity for identifying children at-risk of MH difficulties. However, as is common for predictive validity of screening measures, several studies reported either high specificity or high sensitivity due to the tendency of sensitivity and specificity to be inversely related, and the challenge of developing a measure with maximal precision. Interestingly, the research showed that more educator-report measures had strong predictive validity (Girio-Herrera et al., 2015; Kettler et al., 2017; Ştefan & Miclea, 2017) compared to parent-report (Kettler et al., 2017; Takayanagi et al., 2016). Using the benchmark of at least 70% sensitivity, 70% specificity and 50% PPV for developmental screening measures (Aylward, 1997), the ECS, IRS and PBSS met these standards for educator ratings (Girio-Herrera et al., 2015; Kettler et al., 2017; Ştefan & Miclea, 2017). On the other hand, only the PBSS met this benchmark for parent-report (Kettler et al., 2017). However, results for this measure were mixed, as earlier studies found that the combined, multi-informant predictive validity for the PBSS did not meet standards (Feeney-Kettler et al., 2011).

Surprisingly, the preschool version of the widely-used SDQ demonstrated poor predictive validity, with low sensitivity across both parent and educator report and borderline

acceptable to good specificity for parents and educators (Ezpeleta et al., 2013). Only the hyperactivity subscale had acceptable sensitivity and specificity at the recommended cut points and the study reported poor educator-report screening accuracy in terms of AUC analysis. Other comprehensive reviews of the SDQ as a screening measure amongst preschool populations have found a number of problems with its predictive validity, internal consistency of its subscales, test-retest reliability and criterion (concurrent) validity and have cautioned its use within non-clinical populations (Kersten et al., 2016; Lavigne et al., 2016). The preschool version of the SDQ may not have strong clinical utility or other psychometric properties as it has not been specifically developed for preschool samples, but further research is required to investigate this.

Only one included study reported on the incremental validity of adding educators' ratings to parent ratings, finding a significant improvement in the identification of socio-emotional outcomes over time (Moore, Dowdy, Fleury, et al., 2022). Contrary to previous research with older children, this study found that parent ratings did not significantly improve the prediction of outcomes. Previous findings with older children reported that parent ratings of preschool children added significant variance to child outcomes when considered in addition to later kindergarten teacher ratings (Owens et al., 2015), and a large, longitudinal study of children aged 4–11 years highlighted the power of combining informant reports to predict child outcomes across a number of domains (Verhulst et al., 1994). When evaluating the predictive and incremental validity of these measures, it is important to consider the criterion variables used in each analysis, and the context in which they reference child behaviour, such as the home, education or care setting. When criterion measures are mapped across multiple settings, it is easier to achieve the ideal model of child assessment in which child wellbeing is captured from multiple perspectives (De Los Reyes et al., 2023; Makol et al., 2020). There is promising work emerging which examines the use of screening measures with preschool children through

parent, educator and child psychologists' ratings, highlighting the value of gaining multi-faceted assessments of child wellbeing, but there remains an opportunity to research this further (Gustafsson & Sund Levander, 2024).

These results contribute to the evidence that whilst educators often have minimal MH training, their knowledge of a referent classroom group of children may be advantageous in their assessment of child wellbeing and development, given strong clinical utility for educator-report. Future research should focus on developing and evaluating UMHS measures, with high screening accuracy and clinical utility for use in preschool populations by educators.

Examining the effectiveness of UMHS adds to the weight of evidence for multi-informant screening because, in the context of UMHS, improved identification is clearly linked to increased diagnosis, referral and/or treatment, and thus, improved outcomes for children. Yet our review found no studies reporting on the effectiveness of multi-informant raters. This follows a consistent pattern of omission in the literature in which effectiveness data, in terms of screening outcomes or the consequences of screening, are rarely reported adequately or at all (Houri & Miller, 2020). There is a distinct lack of data reporting screening effectiveness and limited research measuring the effectiveness of UMHS in the child MH space (Brinley et al., 2024). In the case of screening amongst preschool children, none of the included studies investigated the effectiveness of multi-informant screening. This paucity of research remains despite multiple calls that we need to better understand the outcomes and consequences of screening. By effectively understanding outcomes, we can better understand the impact of UMHS (Houri & Miller, 2020). Thus, there is a clear need for future research to examine the effectiveness of screening in preschool populations with additional longitudinal studies measuring MH outcomes.

Given multi-informant report has long been the hallmark of developmental research and despite previous calls to action from highly cited research (e.g., Achenbach et al., 1987;

De Los Reyes et al., 2015), there remains a substantial gap in the study of multi-informant UMHS measures. There is a clear need for more research to demonstrate the incremental value of combining parent and educator reports in UMHS. Given demands on educators' and parents' time, lack of MH training for preschool educators and limited MH resources available to educators, the value of combined contributions needs to be made explicit. By demonstrating the predictive power of multi-informant report, researchers can build the case for UMHS and improve child socio-emotional outcomes.

Critics have raised concerns that UMHS is not warranted in young children as it stigmatises them (Frances, 2012; Jureidini & Raven, 2012), and is neither wanted by parents nor educators (Humphrey & Wigelsworth, 2016). However, there is a mounting case refuting these conjectures. All three studies included in this review that investigated acceptability of UMHS reported strong support from parents and educators for either specific screening measures or UMHS more generally in the preschool-age population. Importantly, each of these respondents evaluated the acceptability of UMHS after having completed screening for their children. A recent review of parent UMHS for children found that parents generally regarded UMHS as appropriate, across a range of contexts including schools, paediatric hospital settings, and well-child visits, although not all studies required parents to actually complete UMHS for their child (Brinley et al., 2024). The findings of the current review, therefore, add to existing research indicating that UMHS is, in fact, considered highly acceptable. Overall, the positive support from parents and educators for screening provides further evidence for the contention that preschools are ideal settings for screening, as they are accessible, less encumbered by MH stigma and a natural fit, given parents frequently turn to educators for MH advice for their children (Desta et al., 2017; Pavuluri et al., 1996).

Whilst the studies in this review largely focussed on the acceptability of specific measures, there is a need to review screening more generally and in the current context. It may

be possible that parents and educators have a growing awareness and understanding of early intervention for supporting child wellbeing, particularly in the aftermath of the global pandemic that had and continues to have significant ramifications on the wellbeing of children. More qualitative and quantitative research should be conducted in this area to identify perceived barriers, facilitators and perceived acceptability of multi-informant screening of preschool-age children.

Strengths and Limitations

This review was the first to identify multi-informant UMHS measures utilised by parents and educators in preschool child populations and the first to evaluate their clinical utility, effectiveness and acceptability. However, there are a number of limitations that need to be kept in mind when interpreting the results. Studies that were excluded from this review included those that were not published in English, did not report predictive validity, effectiveness or acceptability data, or were not peer reviewed. The exclusion criteria for child age may have limited the results and inadvertently excluded studies that included preschool-age populations. It was often difficult to discern if 3–5-year-old children were included in studies with broad-age range samples. The ages of "preschool" children were not always specified and differed between states or countries, as the age of school entry differed across jurisdictions. Secondary source or missing data was not sought from any authors to provide clarification. Moreover, some aspects of the reviewed literature were difficult to synthesise because of the disparate measures utilised. Many studies reported on author-developed measures, with varying identification methods, scoring and subscale outcomes, making synthesis difficult. Finally, risk of bias presents a possible limitation in the area of recruitment bias across the included studies. In evaluating the methodological quality of included studies using the MMAT, some studies may have been biased by not recruiting a representative sample

of the target population (Hong et al., 2018). Given UMHS is intended for use amongst general, non-clinical populations in the community, it is of utmost importance that screening measures are tested in samples which are as representative as possible. The authors suggest that recruitment methodology should match implementation of UMHS in practice in order to minimise confounding variables in testing the efficacy of universal screening of whole populations. For example, UMHS measures should be implemented universally with as few participant eligibility criteria as reasonable and without interference from non-systematic factors such as teacher-nomination.

Implications

Parents and educators of preschool-age children are integral to facilitating help-seeking for young children in need of MH support. They have a central role in early identification of MH problems and are a primary part of the pathway to enabling a child to access healthcare practitioners (De Los Reyes et al., 2015). However, based on the findings of this review, there are few measures that can be recommended to professionals in preschool and early childhood health settings for use with preschool-age children. Based on the clinical utility of the included measures, the PBSS was the only measure to reach benchmark standards for predictive validity for both educator and parent report; however, the measure has not been published or made publicly available (Kettler et al., 2017; Owens et al., 2015). Having demonstrated incremental validity, the patented BASC-3, BESS and the freely available PSC-17 warrant further investigation to determine whether their multi-informant predictive validity are found to be acceptable. It is important to note that the widely used SDQ did not meet acceptable standards for predictive validity and as such, requires further research prior to being used for UMHS purposes in preschool children.

Based on the limited research examining incremental validity of preschool UMHS measures, it would appear that educator ratings are important and add significant incremental value to identifying children where problem behaviour occurs across multiple settings. Given how resource-intensive the implementation of multi-informant UMHS is, there is a strong need for further evidence that examines the incremental validity and effectiveness of UMHS. Future studies that investigate incremental validity need to carefully select gold-standard comparison measures which span various child contexts so that one informant is not inadvertently favoured over others (Haynes & Lench, 2003).

An important implication for future research is the examination of measures in terms of their ease of implementation and use in clinical practice. Some research in the area of measurement-based care has differentiated between the constructs of clinical utility and validity, defining "clinical utility" in terms of ease of implementation and usefulness for clinicians, and "validity" in terms of the evidence for specific measures (McLeod et al., 2022). If adopted in UMHS, this distinction may be of value, as the ease of implementation and usefulness for clinicians and non-clinicians such as educators may be an important factor as to whether UMHS is more widely implemented.

Moreover, there is a need for additional studies examining UMHS, especially those which are representative and broader in scope. Large, population-level studies can provide important normative data that are inclusive of different cultural groups, children in urban and regional contexts and validated across ages. Thus, the results of this review suggest there continues to be a need to examine existing measures and develop new multi-informant UMHS measures for preschool-age children that are psychometrically valid, reliable and accurate for identifying children at risk of MH difficulties. Such research should also enable further examination of acceptability of UMHS in preschool settings.

Conclusion

In examining the available UMHS measures utilised in preschool-age children by parents and educators and by evaluating the psychometric evidence presented in studies, this review found promising results in terms of the clinical utility of measures to identify child MH concerns. However, with few measures reporting on all constructs of interest, there remains a need for further research on clinical utility, effectiveness and acceptability of measures. Accurate identification of young children at risk of MH problems is critical to providing early intervention and disrupting the trajectory of MH problems, and research to identify valid screening measures that can be easily implemented by teachers and educators is an important first step.

CHAPTER 3: Reliability, Predictive Validity and Normative Data for the Pediatric Symptom Checklist-17 (PSC-17) in a National Australian Sample

The contents of this chapter are published as:

McLean, R. K., Tully, L. A., & Dadds, M. (2025). Reliability, predictive validity and normative data for the Pediatric Symptom Checklist-17 (PSC-17) in a national Australian sample. *Australian & New Zealand Journal of Psychiatry*, *59*(8), 702–712. https://doi.org/https://doi.org/10.1177/00048674251342952

3.1 Chapter Synopsis

We evaluated the reliability, validity and acceptability of the PSC-17, a free, brief measure of child MH, in a sample of parents of preschool- (3–5 years) and school-age children (6–17 years). This is the first study to examine parent-reported PSC-17 for children aged three years.

A national community sample of Australian parents (N = 2,097) completed a demographic questionnaire and the PSC-17. We used a cross-sectional, test-retest design to assess the structural validity, internal consistency, test-retest reliability, concurrent and predictive validity of the PSC-17 total and subscale scores. Predictive validity was evaluated using a sub-sample of parents (n = 122) who completed the PSC-17 and CBCL at a second time point. Normative data were also produced.

Factor analysis supported the three-factor model for the PSC-17. Total (r = .820-.932) and subscale scores (r = .748-.891) strongly correlated with the CBCL and demonstrated strong internal consistency (total scores α = .873-.878). Test-retest reliability was acceptable for school-age children. The PSC-17 demonstrated excellent classification accuracy for preschool children; however, it did not perform as strongly in older age ranges. Normative data for age

and gender were produced for the measure. Results indicated high levels of parent acceptability for the measure. Findings contribute new validation evidence for the use of the PSC-17 as a screening and assessment measure in research and clinical settings, for parents of children from three years and above. This study is the first to ascertain Australian normative data for the PSC-17.

3.2 Introduction

Whilst it has been established that many MH problems begin before 14 years of age (Kessler et al., 2007), a recent review of prospective longitudinal studies found MH difficulties experienced in children as young as five years are associated with adult MH disorders (Mulraney et al., 2021). This highlights the need to identify preschool children as a cohort of focus, due to parents' lack of confidence in recognising MH difficulties, the high proportion of children with clinically significant MH problems, and the high levels of unmet MH needs in this age range (Kataoka et al., 2002; Oh et al., 2015; Rhodes, 2017; Robinson et al., 2008).

A plethora of child MH measures currently exist, which may assist parents to identify child MH difficulties. However, there is a lack of available, brief, parent-report measures for screening and assessment that are appropriate for a range of ages, including preschool children aged 3–5 years (McLean et al., 2024; Tully et al., 2024). Measures also need to produce valid and reliable scores (Humphrey & Wigelsworth, 2016). In addition, clinical cutoffs and normative data relevant to specific populations can facilitate the use of measures in research and clinical practice (Kendall & Sheldrick, 2000). The prevalence of child MH problems in Australia is high and stable (Sawyer et al., 2018), despite the increased availability of evidence-based intervention. There is an urgent need to identify measures that can be used to detect children with MH problems in the Australian context. Thus, the aim of this study was to validate and ascertain normative data for the PSC-17 in a national Australian sample.

3.2.1 Overview of the PSC-17

Adapted from a 35-item version (Jellinek & Murphy, 1988), the PSC-17 is a 17-item, parent-report measure of child MH, validated and normed for use with children aged 4–17 years in the US (Gardner et al., 1999). Total scores and three subscales based on internalising, externalising and attention problems are produced by summing responses to items. Items are based on a reflective model and response options include: *Never*, *Sometimes*, *Often*. It has been one of the most frequently recommended screening measures by US MH authorities (Semansky et al., 2003). Within Australia, interest in the PSC-17 has been growing due to the increasing use of digitised measures in clinical and research settings and lack of freely available measures for assessing parent reports of child MH (Tully et al., in press).

Widely used, the PSC-17 is considered feasible to implement in primary care settings due to its advantages over other measures; it is brief, free to implement in digital and paper-based format, easy to comprehend, simple to score, covers a wide range of ages, and is also suitable for use as a screening measure by caregivers who may have limited or no training in MH, such as early childhood educators and teachers (Sheldrick, Henson, Merchant, et al., 2012).

3.2.2 Psychometric Properties

The technical accuracy of measures, as indicated by various psychometric or measurement properties, is critical when considering the use of measures in screening and assessment of child MH. The PSC-17 has established validity and reliability in children aged 4–17 years and, as part of a recent review of 95 child MH measures, was rated as having excellent psychometric properties in terms of reliability, validity, normative data and treatment

sensitivity; one of only 21 measures to be rated as such (Becker-Haimes et al., 2020; Spencer et al., 2020; Stoppelbein et al., 2012). Multiple studies have reported strong internal consistency for the PSC-17; Cronbach's alpha .85–.89 for total scores, .78–.79 internalising, .80–.83 externalising and .82–.83 attention (Gardner et al., 1999; Murphy et al., 2016; Wagner et al., 2015).

There is limited research on the test-retest reliability of the PSC-17. One community study reported good test-retest reliability (intraclass coefficient [ICC] = .85) after 8–14 days (Murphy et al., 2016). Another study reported an ICC of .55; however, the retest was implemented at six months amongst children in foster care (Jacobson et al., 2019).

Previous research has reported the PSC-17 subscales had significant relationships with associated diagnosed disorders; e.g., internalising subscale scores with depression and anxiety, indicating construct and convergent validity (Jacobson et al., 2019). The original validation study for the PSC-17 found total scores correlated with the clinician-rated Children's Global Assessment Scale (CGAS), r = -0.64; 95% CI -0.71 to -0.55, and CBCL, r = -0.60; 95% CI -0.67 to -0.52 (Gardner et al., 2007).

The PSC-17 has strong predictive validity, which is important for screening measures. A systematic review of screening measures identified three studies reporting high sensitivity (SE) .82-.95 and specificity (SP) .81-.91 (Lavigne et al., 2016). Mean sensitivity was .90 and specificity .85, which surpassed benchmark standards for developmental screening measures (Aylward, 1997; Lavigne et al., 2016).

Whilst the PSC-17 has been well-validated in older children, there remains a need to evaluate the PSC-17 as a measure of child MH in young children in terms of factor structure, internal consistency, test-retest reliability, concurrent and predictive validity.

3.2.3 Use of the PSC-17 with Preschool-age Children

While much of the existing research on the PSC-17 has been on school-age children, limited research has examined the predictive validity and reliability of the PSC-17 in preschools. Studies with teacher report found strong internal consistency across total scores and subscales (DiStefano et al., 2017; Liu et al., 2020; Moore, Dowdy, Fleury, et al., 2022). In a large paired-sample study, partial scalar invariance was established, indicating parents and teachers of children aged 4–5 years interpreted items similarly (Gao et al., 2022). However, there remains a need to validate parent-reported ratings in children aged three years, since the measure is currently being used in clinical practice with this age (Wagner et al., 2015; Benheim, T., Jellinek, M. & Murphy, M., personal communication, March 7, 2023).

3.2.4 Normative Data

The PSC-17 has not been validated with parents in Australia, nor is there normative data outside of the US, which is important for the systematic interpretation of results and efficient identification of children who may have emerging MH problems (Sheldrick, Henson, Merchant, et al., 2012). Moreover, the PSC-17 does not have normative data for the preschool age in any country. Using normative data developed for older children in clinical settings may not be appropriate for young children in community settings. US data also may not be appropriate in the Australian cultural context. Producing localised normative data ensures recommended clinical cut-offs are appropriate for screening and assessment purposes in non-clinical populations.

3.2.5 Acceptability

Some researchers and practitioners have considered the use of MH screening measures with young children as contentious (Frances, 2012; Jureidini & Raven, 2012). The (lack of) acceptability of a measure can affect uptake and implementation of child MH or screening measures across settings (Harrison et al., 2013). Acceptability may be defined by its perception as helpful or useful, whether it is easy to understand and whether users would recommend it to others (Brinley et al., 2024; Humphrey & Wigelsworth, 2016). Therefore, the perceived acceptability of a measure by users (in this case parents) is an important aspect to investigate if it is to be adopted as part of improving access to early intervention.

3.2.6 The Current Study

This study seeks to add to existing evidence by evaluating the parent-reported PSC-17 and ascertaining normative data across preschool and school-age populations. To our knowledge, this is the first study to validate the PSC-17 for children aged three years and the first to ascertain normative data, including age and gender versions for an Australian population. The aims of the study were, first, to evaluate the psychometric properties of the PSC-17 and validate parent ratings for preschool children (3–5 years) and school-age children (6–17 years), in terms of factor structure, internal consistency, test-retest reliability, and concurrent and predictive validity. Second, to ascertain the specific age and gender normative data for parent-report PSC-17 in Australia for children aged 3–17 years, and finally, to examine the acceptability amongst parents of preschool children compared to school children.

3.3 Method

3.3.1 Procedures

Study approval was obtained from the University of Sydney Human Research Ethics Committee. There are no set conventions specific to sample sizes for establishing normative data, albeit ideal samples should be large and representative of the defined population. Previous normative studies have recruited samples varying from 500 to 5,400 participants (D'Souza et al., 2017; Mellor, 2005). Based on methods previously used by the investigators, this study recruited two samples of 1,000 parents of preschool (3–5 years) and school-age children (6–17 years).

Participants were recruited through an online research panel and sampled to represent Australian families based on key socio-demographic characteristics from national census data (i.e., household income, marital status, and residential location). Inclusion criteria included adult parents or caregivers of a child aged 3–17 years who reside in Australia and have basic English literacy. Caregivers could include anyone in a caregiving role such as fathers, mothers, kinship carers, or foster carers. Quotas were used for the recruitment of two samples, including an even ratio of child gender and ages within each sample. See Appendix A-3 for further details about recruitment procedures.

After providing informed consent, participants completed measures which were counterbalanced in order of completion. Baseline questionnaires took approximately 10–15 minutes to complete. After data quality checks, the final sample included parent reports of 1,048 preschool and 1,049 school-age children.

To evaluate retest-reliability, concurrent and predictive validity, a sub-sample of participants completed a second set of questionnaires, which took approximately 15–20 minutes to complete. A subset of preschool parents (n = 55) completed questionnaires 6–73 days after baseline (M = 35) and 67 school parents 11–69 days (M = 43) later. In the preschool

sample, 16.4% of parents, and in the school sample, 4.5% of participants completed the second assessment between 7–14 days after baseline.

3.3.2 Participants

Preschool children were aged 3.87 years on average (*SD* = .78) and school children were aged 11.28 years on average (*SD* = 3.42). Preschool and school-age samples were almost evenly split between male and female (see Table 4). One preschool child and four school-age children were identified as non-binary gender. Child ethnicity was predominantly Caucasian, followed by Aboriginal and Torres Strait Islander and multiple ethnicities (see Table 5). Parents from all states and territories were represented in both samples and were culturally diverse. See Appendix A-4 for further details.

3.3.3 Measures

In addition to the PSC-17, parents completed questions about child and family socio-demographic details and the following measures.

Child Behavior Checklist

The CBCL (Achenbach, 1999) is a comprehensive measure of child MH with strong psychometric properties across two age versions – 1.5–5 years and 6–18 years. It is the most common criterion measure in classification accuracy studies (Lavigne et al., 2016). The CBCL consists of 99 items (1.5–5 years) and 113 items (6–18 years) and produces internalising, externalising and total problems scales. The CBCL was used in the current study as a criterion measure to establish concurrent and predictive validity.

Table 4.

Age and Gender Distribution of the Sample

Child age in years	Male N	Female N	Total N
3	209	186	395
4	205	187	393*
5	141	119	260
6	42	46	88
7	55	48	103
8	44	50	94
9	48	43	91
10	46	43	89
11	40	47	87
12	48	38	86
13	44	49	93
14	44	40	85*
15	29	45	76*
16	39	39	79*
17	35	43	78
Total	514	531	2,097

Note. *Five children were identified as non-binary gender.

Table 5.

Child Ethnicity for PSC-17 Sample

Child ethnicity	Children aged 3–5 years		Children aged 6–17 years	
	N	Percent	*N*	Percent
Aboriginal/Torres Strait Islander	76	7.2	62	5.9
African/African American	5	0.5	8	0.8
Caucasian (e.g., British, European)	685	65.3	762	72.6
East Asian (e.g., Chinese, Japanese, Korean)	42	4.0	29	2.8
Hispanic/Latino	3	0.3	4	0.4
Middle-Eastern (e.g., Egyptian, Iraqi, Lebanese)	10	1.0	11	1
Polynesian (e.g., Pacific Islander, Māori)	12	1.1	6	0.6
South Asian (e.g., Indian, Pakistani)	53	5.0	37	3.5
South-East Asian (e.g., Vietnamese, Filipino)	45	4.3	24	2.3
Other or multiple ethnicities identified	117	11.2	106	10.1
Total	1048	100	1049	100

Acceptability

Parents' views on the acceptability of the PSC-17 were assessed using four items adapted from a parent-report measure developed by Hawes et al. (2021). Responses were collected on a 5-point Likert response scale from *not at all/strongly disagree* (1) to *very/strongly agree* (5). See Appendix A-5 for further details about this measure.

Analytic plan

Factor structure, reliability and validity were evaluated separately for preschool and school-age children. Using the original three-factor model (Gardner et al., 1999), Confirmatory Factor Analysis (CFA) was performed and fit statistics reported. CFA was conducted using Amos (IBM SPSS Statistics, Version 29.0.0). The latent variables, internalising, externalising and attention, were allowed to be correlated, and all measurement error was presumed to be uncorrelated. Hu and Bentler's (1999) standards of acceptable model fit were used to evaluate the model. Further details are outlined in Appendix A-6.

To align with the age of school entry (preschool, primary and secondary school) and CBCL scoring, analyses for predictive validity were conducted on young children (3–5 years), older children (6–11 years), and adolescents (12–17 years). Receiver operator characteristic (ROC) analysis in each of these age groups examined the association between PSC-17 subscales and CBCL case classifications (clinical/non-clinical cases). Normative data is also presented for these age ranges. As the online questionnaire required input for all items, there was no missing data.

Since the test-retest time interval was longer than intended, reliability estimates for short intervals (7–14 days) compared to those outside that time period (< 7 and > 14 days) were analysed, and no significant difference was found. Thus we report retest reliability for the full sub-sample.

Psychometric properties were evaluated against Youngstrom and colleagues' criteria (2017) set out in a rubric by De Los Reyes and Langer (2018). Since the rubric does not include specific thresholds for predictive validity or classification accuracy, predictive validity of the scale was evaluated against benchmark criteria for developmental screening measures of 70% and above for sensitivity and specificity (Aylward, 1997) and Area Under Curve (AUC) values of .80 or higher were considered "very good" and .90 or higher as "excellent" (Chaffin et al., 2017).

Normative data was determined by reporting scoring cut-offs for the 90th (borderline) and 95th percentile (at-risk cases) for subscales and total scores, gender and age. At-risk scores indicate positive identification on the screening measure and further investigation of child MH may be warranted.

3.4 Results

Factor structure

For the preschool sample, the chi-square fit statistic was large and significant, χ^2 (116, N = 1,048) = 661.735, $p < .001$. χ^2/df = 5.70. Other fit statistics, however, revealed an RMSEA = .07, CI 90% .062–.072, SRMR = .0530, CFI = .91, TLI = .898, indicating some support for the hypothesised model. Factor loading estimates were moderately to strongly correlated to the latent factors (R^2 0.49–0.82) in the preschool sample.

For the school sample, the chi-square fit statistic was large and significant, χ^2 (116, N = 1,049) = 474.837, $p < .001$. χ^2/df = 4.09. Other fit indices supported the hypothesised model in the school-age sample, RMSEA = .054, CI 90% .049–.059, SRMR = .0469, CFI = .943, TLI = .933. Factor loading estimates were moderately to strongly correlated to the latent factors (R^2 0.57–0.84) in the school sample.

The modified preschool model and the school model met three out of the six criteria, with two criteria “close to” cut-offs.

Distribution of scores

The mean, standard deviation and ranges for each of the samples are outlined in Table 6.

Reliability

In the preschool sample, good internal consistency was found for the PSC-17 total scores (α = .87), externalising (α = .83), and attention subscales (α = .84). Internal consistency of the internalising subscale was adequate (α = .73). In the school sample, internal consistency for total scores (α = .88), externalising (α = .83), and attention subscales (α = .84) was also good. Internal consistency of the internalising subscale was adequate (α = .78). .

In the preschool sample, analyses for test-retest reliability were performed and reported as follows: total scores (ICC =.66), internalising (ICC = .62), externalising (ICC = .61), and attention subscales (ICC = .63). In the school sample, good test-retest reliability was found for total scores (ICC =.73), and the internalising subscale (ICC = .70).Test-retest reliability for externalising (ICC = .67) and attention subscales (ICC = .68) did not meet adequate criteria. ICC estimates and their 95% confident intervals were calculated based on a single rating, absolute-agreement, 2-way mixed-effects model.

Table 6.

Mean Scores and Banding for PSC-17 by Age and Gender

		3–5 years (n = 1,048)	6–17 years (n = 1,049)
Internalising	*M*	1.95	3.50
	SD	1.80	2.19
	Range	0–10	0–10
	Possible range	0–10	0–10
	α	.73	.78
	Top 5% of cases	≥5	≥7
	Top 10% of cases	≥4	≥6
Externalising	*M*	3.66	3.41
	SD	2.86	2.96
	Range	0–14	0–14
	Possible range	0–14	0–14
	α	.83	.83
	Top 5% of cases	≥9	≥9
	Top 10% of cases	≥7	≥7
Attention	*M*	3.12	3.68
	SD	2.59	2.78
	Range	0–10	0–10
	Possible range	0–10	0–10
	α	.84	.84
	Top 5% of cases	≥8	≥9
	Top 10% of cases	≥7	≥8
Total scores	*M*	8.73	10.59
	SD	6.04	6.30
	Range	0–34	0–32
	Possible range	0–34	0–34
	α	.87	.88
	Top 5% of cases	≥20	≥22
	Top 10% of cases	≥17	≥19

Concurrent validity

The PSC-17 strongly correlated with the CBCL across ages, demonstrating concurrent validity. In the preschool sample, PSC-17 total scores strongly and significantly correlated with the CBCL total problems at the bivariate level, $r = .93$, $p < .001$, $n = 55$. PSC-17 subscales also strongly correlated with the CBCL subscales: internalising $r = .86$, $p < .001$, externalising $r = .86$, $p < .001$; and attention $r = .89$, $p < .001$. In the school sample, PSC-17 total scores were strongly associated with the CBCL total problems, $r = .82$, $p < .001$, $n = 67$; internalising $r = .75$, $p < .001$; externalising $r = .80$, $p < .001$; and attention $r = .81$, $p < .001$.

In the preschool sample, the AUC for PSC-17 total scores was excellent, .86 with a standard error of .12. This was significant at .002 (CI .628–1.084). Internalising AUC was adequate .68, SE .14, p =.184 (CI .414–.947); externalising AUC was excellent .83, SE .14, p = .018 (CI .555–1.101); and attention AUC was excellent .89, SE .05, $p < .001$ (CI .792–.996). In the school sample, the AUC for PSC-17 total scores was excellent, .83 with a standard error of .07. This was significant at $p < .001$ (CI .683–.967). Internalising AUC .79, SE .07, $p < .001$ (CI .649–.926); externalising AUC .88, SE .08, $p < .001$ (CI .713–1.038); and attention AUC .91, SE .08, $p < .001$ (CI .758–1.065) subscales all met criteria for excellent validity.

Predictive validity

Predictive validity was examined in young children (3–5 years, $n = 55$), older children (6–11 years, $n = 30$), and adolescents (12–17 years, $n = 37$). A low number of clinical cases across ages meant assumptions were violated for Chi-Square tests. The older children and adolescent categories were combined for analysis; however, case numbers for the attention subscale remained low. In the preschool sample, high sensitivity was demonstrated for total scores (83.3), externalising (83.3) and attention subscales (77.8). Very high specificity was demonstrated for total scores and all subscales (93.3–97.8). In the school sample, moderate

sensitivity was demonstrated for the externalising subscale only (71.4). High specificity was demonstrated for total scores and all subscales (88.7–100). Predictive validity of the PSC-17 using original clinical cut-off scores for each of the samples is presented in Table 7. ROC analysis identified alternate cut-off points which maximised sensitivity and specificity. Alternate clinical cut-points and their respective sensitivity and specificity are presented in Table 8.

Normative data

In the preschool-age range, 17% of cases were positively identified (at-risk) on total scores; 9.4% internalising; 16.5% externalising; and 13.5% attention. In the school-age range (6–17 years), 25.3% were positively identified on total scores; 31.8% internalising; 15.1% externalising; and 19.2% attention. Fine-grained age analysis revealed 25.4% of older children were positively identified on total scores; 26.1% internalising; 16.8% externalising; and 20.5% attention. Amongst adolescents, 25.2% were positively identified on total scores; 38.2% internalising; 13.1% externalising; and 17.7% attention. For each age and gender group, table 9 presents cut-off scores for the highest-ranking cases.

Males across ages scored higher than females on total scores; as such, further analyses were performed to test the significances of these differences by gender. Preschool males (M = 9.49) scored significantly higher than females (M = 7.89) on total scores, as indicated by an independent samples t-test (F = 7.79, $p < .001$). A two-way ANOVA indicated there was no significant interaction between the effects of preschool age and gender on PSC-17 total scores. Whilst total scores were also higher for school males (M = 11.04) than for females (M = 10.10), this was not a significant difference (F = 2.98, p = .085).

Table 7.

Predictive Validity of PSC-17 using Original Clinical Cut-off Scores

PSC-17	Cut-off score	AUC	Sensitivity %	Specificity %	PPV	NPV
3–5 years (*n* = 54)						
Internalising	5	.80 (CI .63–.98)	25	97.8	66.7	73.7
Externalising	7	.82 (CI .55–1.10)	83.3	95.8	71.4	97.9
Attention	7	.87 (CI .74–.99)	77.8	93.3	70	95.5
Total score	15	.86 (CI .63–1.08)	83.3	93.8	62.5	97.8
6–17 years (*n* = 67)						
Internalising	5	.79 (CI .65–.93)	0	100	0	77.3
Externalising	7	.88 (CI .71–1.04)	71.4	91.5	50.0	96.4
Attention	7	.91 (CI .76–1.07)	0	100	0	92.4
Total score	15	.83 (CI .68–.97)	61.5	88.7	57.1	75.8

Table 8.

Predictive Validity of PSC-17 using Alternate Clinical Cut-off Scores

PSC-17	Cut-off score	Sensitivity %	Specificity %
3–5 years (*n* = 54)			
Internalising	2.5	75	76.1
Externalising	6.5	83.3	95.8
Attention	3.5 or 7.5	100 or 66.7	58.3 or 91.7
Total score	17	83.3	95.8
6–11 years (*n* = 30)			
Internalising	5	40	80
Externalising	0.5	50	92.3
Attention	0.5	100	92.6
Total score	16.5	60	96
12–17 years (*n* = 37)			
Internalising	4.5	80	84.6
Externalising	6	100	90.9
Attention	2.5	100	55.9
Total score	11.5	87.5	85.7

Table 9.

Mean Scores and Banding for PSC-17 by Age and Gender (Subgroups)

		3–5 years		6–11 years		12–17 years	
		Male (*n* = 555)	Female (*n* = 492)	Male (*n* = 275)	Female (*n* = 277)	Male (*n* = 239)	Female (*n* = 254)
Internalising	*M*	2.04	1.85	3.46	3.01	3.64	3.86
	SD	1.87	1.71	2.09	1.76	2.47	2.31
	Range	0–10	0–9	0–10	0–9	0–10	0–10
	α	.75	.71	.76	.67	.83	.80
	PSC case rate (% positive)	10.1	8.7	31.3	20.9	36.8	39.0
	At-risk (top 5%)	6	5	7	6	8	8.25
	Borderline (top 10%)	5	4	6	5	7	7
Externalising	*M*	3.98	3.30	3.96	3.41	3.20	2.96
	SD	3.03	2.62	3.11	2.90	2.83	2.87
	Range	0–14	0–14	0–13	0–14	0–14	0–14
	α	.84	.80	.84	.83	.80	.83
	PSC case rate (% positive)	20.5	12.0	21.8	11.9	12.6	13.0
	At-risk (top 5%)	10	8	10	9	8	8
	Borderline (top 10%)	8	7	8	7	7	7
Attention	*M*	3.48	2.73	4.20	3.58	3.53	3.38
	SD	2.70	2.41	2.75	2.81	2.73	2.75
	Range	0–10	0–10	0–10	0–10	0–10	0–10
	α	.85	.82	.84	.85	.84	.84
	PSC case rate (% positive)	16.9	9.6	23.3	17.7	18.0	17.7
	At-risk (top 5%)	9	8	9	9	9	9
	Borderline (top 10%)	8	6	8	8	8	8
Total scores	*M*	9.49	7.89	11.62	10.01	10.37	10.20
	SD	6.41	5.47	6.51	5.82	6.48	6.26
	Range	0–34	0–29	0–32	0–28	0–30	0–31
	α	.90	.87	.89	.86	.89	.88
	PSC case rate (% positive)	20.5	13.0	29.5	21.3	25.1	24.4
	At-risk (top 5%)	22	18	24	22	23	21
	Borderline (top 10%)	18	16	21	19	19	19

Note. Data not reported for non-binary gender due to the low number of cases.

Acceptability

Parents' ratings of dimensions related to the acceptability indicated that most parents found the PSC-17 to be very or quite appropriate (preschool 87.3%, *n* = 55; school-age 82.9%, *n* = 1,049) and very or quite clear (preschool 92.8%; school-age 93.3%). Most parents were very or quite confident in the ratings they provided (preschool 96.3%; school-age 95.3%); and agreed or strongly agreed that the PSC-17 asked important questions about their child's MH (preschool 89.1%; school-age 92.5%). Independent t-tests showed that there was no significant difference between ratings of acceptability between preschool parents and school-age parents.

3.5 Discussion

The purpose of this study was to evaluate the reliability and validity of the parent-reported PSC-17 in young children, ascertain Australian normative data and examine acceptability amongst parents. To address the first aim, this study presents new evidence in support of the measure's psychometric properties and validates parent ratings for children aged 3–17 years, in terms of factor structure, reliability and validity. Considered alongside the factor loadings, the three-factor model was supported across preschool and school-age ranges, which is consistent with previous studies which included 4–15-year-old children (DiStefano et al., 2017; Murphy et al., 2016).

Overall, using specified evaluation criteria, the PSC-17 for preschool parents was found to have good internal consistency, excellent concurrent validity and good predictive validity. For school parents the PSC-17 was found to have good internal consistency, good test-retest reliability and excellent concurrent validity. In terms of reliability, the PSC-17 demonstrated strong internal consistency across ages. Test-retest reliability fell just below acceptable levels for the preschool sample according to the specified evaluation criteria; however, this analysis was hampered by a large variation and lengthy time period between testing. Given the rapid

child development that occurs during the preschool years, it is not surprising that preschool ICCs were lower than school-age coefficients. The acceptable test-retest reliability for school-age children was higher than other studies which had longer, 6-month retest periods (ICC = .55) (Jacobson et al., 2019) and on par with more ideal 8–14 day retest period (ICC = .85) (Murphy et al., 2016).

In terms of concurrent validity, the PSC-17 correlated strongly with the CBCL across subscales and total scores for both samples. Moreover, ROC analyses revealed the PSC-17 had excellent classification accuracy with very high AUC results across ages. The AUC for the internalising subscale for the preschool sample was borderline acceptable and acceptable for the school sample, which was lower than the AUC of 82% found in the original derivation study (Gardner et al., 1999). That study included a large sample of clinical and primary care presentations, which may explain the difference.

In terms of predictive validity, the PSC-17 demonstrated excellent screening accuracy for total scores, externalising and attention subscales using the recommended cut-offs for preschool-age children. Surprisingly, the PSC-17 did not perform as strongly in the older age ranges. Whilst excellent specificity was found for all subscales and total scores for school-age children and positive predictive values (PPV) were acceptable for externalising and total scores, only the externalising subscale passed benchmark standards for combined high sensitivity, specificity and PPV. Results may have been hampered by a low number of clinical cases, as is common when recruiting from community samples. Further work is required to investigate sensitivity for school-age children, since numerous studies have attained high sensitivity and high specificity for the PSC-17 in this age range (Lavigne et al., 2016).

Normative data

The second aim of this study was to ascertain age and gender normative data for parent-report PSC-17 in Australia for children aged 3–17 years. This study found positive screening rates of 17% for preschool children and 25.3% for school-age children, which is higher than the most recent Australian child and adolescent survey of MH, which identified 13.9% of children aged between 4–17 years as having a mental disorder (Lawrence et al., 2015), and higher compared to the two previous normative studies from the US, which were reported as 15% for children of 4–15 years (Gardner et al., 1999) and 11.6% (Murphy et al., 2016).

The higher number of cases may be due to differences in the local population and timing of data collection. It has been established over the past decade that rates of child MH in Australia are high and stable (Sawyer et al., 2018), and Australian-born children are more likely to be positively identified on screening and have mental illness compared to children born elsewhere (Australian Institute of Health and Welfare (AIHW), 2020). Cut-offs on similar measures are also higher in Australia compared to samples in the United Kingdom and US (Kremer et al., 2015).

Elevated scores of child MH may also be due to the effects of the COVID-19 pandemic on Australian children, who experienced a national lockdown, restricted access to schooling and social supports (Sicouri et al., 2023). Research has shown Australian parents had poorer MH and functioning indicating the pandemic had a significant, negative impact on Australian families (Westrupp et al., 2023). Whilst robust systematic comparisons of current child MH symptoms to pre-pandemic data is still emerging, this study's results suggest further investigation is required into the ostensibly high rates of children identified with emerging MH difficulties and is a consideration for the generalisability of the results.

Acceptability

The final aim of this study was to examine acceptability. Results indicated that parents found the PSC-17 to be highly appropriate, easy to understand and important in terms of assessing their child's MH, and there was no difference in levels of acceptability between preschool and school-age parents. The construct of acceptability is associated with the likelihood that screening measures are adopted in future and thus, these results are a promising indication of parental support for the use of this measure (Glover & Albers, 2007; Kamphaus et al., 2007). High levels of acceptability may be associated with an increased parental awareness of child MH challenges in the aftermath of the pandemic.

Strengths and Limitations

Strengths of this study included the large and diverse sample recruited from across Australia and the use of a comprehensive, comparative measure in the form of the CBCL. However, the representativeness of the sample cannot be determined beyond comparative analysis to census data. Moreover, the CBCL is not a gold-standard, criterion measure nor is it a diagnostic measure. Furthermore, recruiting from the community combined with a small sub-sample at the second time point meant that a limited number of clinical cases were identified. This affected the predictive validity analysis; however, low clinical cases are symptomatic of recruiting from a non-clinical population. Common method variance may have also affected results, since both assessment measures were questionnaires. Future studies should seek to replicate this study's findings and employ alternate methods of assessment, such as comprehensive clinical interviews to assess diagnostic and screening accuracy. Further, future research may investigate the measure's responsiveness and sensitivity to change, which are important measurement properties in clinical and research contexts.

Implications

The findings of this study support the PSC-17 as a valid, reliable and acceptable parent measure for preschool- and school-age children within this population. As a freely available, brief measure, it can be adopted widely in a range of contexts. By spanning the ages of 3–17 years, it enables measurement continuity, that is, using the same measure at different ages, which is particularly important for longitudinal research, in which children often age out of measures and cross-measure comparisons are difficult. Continuity of measures may also benefit clinical work, such as measurement-based care or universal screening conducted at discrete time points in a child's life, e.g., at preschool, school entry, or high school graduation. In fact, recent surveys have found practitioners perceive brief and free measures as important facilitators to increase the use of measures in practice (Tully et al., 2024). Thus, the PSC-17 may serve as an important instrument for tracking the progress of clinical interventions to improve child MH.

Conclusion

The current study contributes new evidence for the PSC-17 as an appropriate and acceptable measure of child MH, which can be utilised in clinical and community settings for a wide range of ages including preschool-age children. This study was the first to ascertain normative data for the PSC-17 outside of the US and for children aged three years. By ascertaining normative data for the Australian population and young children, children with emerging MH difficulties can be accurately identified and directed towards further assessment and early intervention.

CHAPTER 4: Multi-informant Reports of Preschool Mental Health: Validation of Parent and Educator Reports and Normative Data for the Preschool Pediatric Symptom Checklist and PSC-17

The contents of this chapter are published as:

McLean, R. K., Tully, L. A., & Dadds, M. R. (2025). Multi-informant reports of preschool mental health: Validation of parent and educator reports and normative data for the preschool Pediatric Symptom Checklist and PSC-17. *Child and Adolescent Psychiatry and Mental Health*, *19*(1), 138. https://doi.org/10.1186/s13034-025-00985-3

4.1 Chapter Synopsis

The prevalence of MH problems and unmet needs in preschool-age children highlights the challenge of identifying emerging difficulties using validated measures. Given the lack of existing, brief screening measures for preschool-age children, especially multi-informant measures, this study examined two versions of the PSC as reported by parents and educators, the Preschool PSC and the PSC-17. In line with Standards for Reporting Diagnostic accuracy studies (STARD) guidelines, this study examined the psychometric properties, scoring thresholds, and acceptability of parent-reported Preschool PSC. It also examined the psychometric properties of educator ratings for both PSC measures and whether educator ratings improved the incremental validity of educator and parent ratings. Two studies present validation evidence for two MH measures for use with preschool-age children. Participants were a nationally representative sample of Australian parents (n = 1,045; study 1) and a paired sample of parents and educators (n = 94 dyads; study 2) of children aged 3–5 years.

Results supported the internal consistency, test-retest reliability and concurrent validity of the Preschool PSC. Parents and educators indicated high levels of acceptability of both PSC

measures. Results indicated parent-reported Preschool PSC and PSC-17 significantly improved the prediction of clinician-rated functioning scores over and above educator report suggesting incremental validity for multi-informant report. Normative data for the parent-reported Preschool PSC are presented for the first time. This research expands the evidence base for the validity, reliability and acceptability of the parent- and educator-report Preschool PSC and PSC-17 measures as utilised with young children. Although further research is required, this research contributes new evidence, including incremental validity and normative data, to increase the clinical utility of both PSC measures.

4.2 Introduction

Despite the early onset and prevalence of MH problems in preschool children aged under five years (Mulraney et al., 2021), young children are less likely to access services for their MH concerns compared to adolescents (Lawrence et al., 2015). Prevalence estimates of child MH problems in preschool-aged children vary widely from 7.1–26.4% (Egger & Angold, 2006; Gudmundsson et al., 2013; Oh et al., 2015) and without access to services or treatment, problems can remain stable, persist into adulthood and have long-term negative ramifications (Caspi et al., 1996; Jones et al., 2015; Klein et al., 2015). Early identification of MH problems can improve outcomes (Lipkin & Macias, 2020); however, in order to identify children in need of intervention, validated measures suitable for use in the preschool population, using multi-informant reports from primary caregivers, are needed. The Preschool PSC and PSC-17 are broad-based child MH measures, but there are significant gaps in the research supporting these measures with parents and educators of preschool-aged children; they require further validation and normative data in order to have greater clinical utility and be adopted in research and clinical practice. This study presents new validation evidence for the Preschool PSC and PSC-

17 for parent and educator reports for preschool-age children and provides normative data for the Preschool PSC for the first time.

4.2.1 Universal Mental Health Screening

The prevalence of problems in preschool-age children highlights a need to identify MH difficulties early so that children can receive intervention. One tool for early identification is a population-based approach called universal screening, which is also known as systematic monitoring, routine surveillance, or wellbeing check-ins. Screening involves the use of standardised measures, such as the Patient Health Questionnaire (Spitzer et al., 1999) and Spence Children's Anxiety Scale (Spence & Rapee, 1999) to identify children who may be at risk of MH problems. The measures are rated by single or multiple informants, and generally quick and inexpensive to implement (Albers et al., 2007). Screening is particularly important in the early years when MH problems are less entrenched and severe (Humphrey & Wigelsworth, 2016).

Screening measures or check-ins can be useful for parents, health professionals, and educators. Completion of wellbeing check-ins may lead to an opportunity to discuss child MH, since research suggests few parents or educators will raise concerns with each other spontaneously and instead, will wait for other caregivers to alert them to potential concerns (Burkett et al., 2025; Honda et al., 2023). This emphasises the importance of developing multi-informant tools that are considered acceptable by parents and educators.

4.2.2 Multi-informant Screening Utilising Parents and Educators

Whilst reports between raters are often not highly consistent, multi-informant assessment is a hallmark of developmental psychology and essential for early identification

(Barbarin et al., 2020; De Los Reyes et al., 2015). Multi-informant measures are needed to identify children at risk of MH problems; however, two recent systematic reviews have highlighted there are few measures validated for use with preschool-aged children either with single or multiple raters (Becker-Haimes et al., 2020; McLean et al., 2024). In order to allow multi-informant reports, there is a need, therefore, to establish psychometrically sound measures for young children that can be used with parents and educators.

Early childhood educators' perspectives of child MH are valuable, as they observe children in naturalistic play for many hours and can compare child MH with a referent group of peers (Franck et al., 2024; Stormont & Stebbins, 2005). Whilst medical professionals are increasingly using screening measures in their practice (Brener & Wright, 2025), educators often lack MH training and knowledge around measure use (Reinke et al., 2011). Yet, educator reports are one of the primary pathways of identification of MH problems in young children (Steed & Banerjee, 2016) and educators are a primary source of advice when parents are help seeking for their child (Pavuluri et al., 1996). There is, therefore, a need to provide parents and educators with measures that will help them identify children at risk of MH difficulty.

Whilst adding educator perspectives to parent-report measures has the potential to increase the validity of screening, it can also raise the possibility of discrepant informant ratings (De Los Reyes et al., 2015). Although discrepancies between informants may appear disadvantageous, they can in fact offer important, domain-specific information about a child's behaviour informed by the informant's context (De Los Reyes et al., 2023). When multiple perspectives about a child are captured from various contexts like the home and preschool, the ideal model of child assessment is more likely to be obtained (Makol et al., 2020). Moreover, incremental validity is important for multi-informant measures, as it provides an indication of how much each informant's perspective contributes to a specified outcome and whether additional perspectives are justified. However, incremental validity of multi-informant

measures for preschool children has not been well researched. A recent systematic review identified only one study investigating incremental validity of two measures; a significant improvement in identification of child MH problems was found when educator ratings were added to parent ratings for the PSC-17 and Behavior Assessment System for Children-3 Behavioral and Emotional Screening System (McLean et al., 2024). The lack of research in this area reveals a need to understand the incremental value of collecting data from multiple caregivers.

4.2.3 The Pediatric Symptom Checklists

Whilst there are several MH screening measures which have been identified in previous reviews of the literature, only a few have been validated for use with preschool-age children, and these measures have varying psychometric strength (Becker-Haimes et al., 2020; Lavigne et al., 2016). Psychometric properties such as internal consistency, test-retest reliability, inter-rater reliability, and predictive validity are important for ensuring measures are free from measurement error and are suitable for use with the population in which they were tested. Two Pediatric Symptom Checklists (PSCs) have been recommended for use with preschool populations: the PSC-17 and Preschool PSC. Both are brief, free in digital and paper-based formats, and easy to score. They produce total scores and subscales based on internalising, externalising and attention problems. The PSC-17, which is widely used, was developed for older children, but has recently been validated for use with children aged 3–5 years (McLean et al., 2025). However, investigations of educator report for the PSC-17 are limited (Gao et al., 2022; Liu et al., 2020; Moore, Dowdy, Fleury, et al., 2022). Previous studies have investigated the factor structure of educator-rated PSC-17, inter-rater reliability and incremental validity in the preschool environment (DiStefano et al., 2017; Liu et al., 2020; Moore, Dowdy, Fleury, et al., 2022); however, other facets of reliability and validity of educator report with a paired

sample of parents and educators have not been investigated for the PSC-17. Further, the convergence of educator report with diagnostic clinical interviews has not been investigated for this measure. The 18-item Preschool PSC is a related measure, adapted from the PSC-17 and specially developed for preschool-aged children 18 months–5.5 years, with developmentally appropriate item wording and a fourth subscale called "Parenting Challenges", which assesses parent difficulties across six items. However, the developers of this scale caution against interpreting any of the subscales, due to concerns over reliability, and instead suggest use of the total score only. Whilst previous research on the Preschool PSC features diverse samples (Leiva et al., 2024; Suman et al., 2022), there is no normative data and the measure has not yet been examined with educators, limiting its clinical utility. The Preschool PSC has strong internal consistency, Cronbach's alpha .80–.92, and good test-retest reliability (ICC = .75) after four weeks (Rocha & Nunes, 2021; Sheldrick, Henson, Merchant, et al., 2012). Classification accuracy or predictive validity is important for screening measures; the Preschool PSC has high sensitivity (.87–1.00) when tested against the CBCL and the Ages & Stages Questionnaires (Sheldrick, Henson, Merchant, et al., 2012). Recent research reported the Preschool PSC has moderate predictive validity of sensitivity .75; specificity .77; Positive Predictive Value (PPV) .60; Negative Predictive Value (NPV) .99, compared to International Classification of Disease scores after 1 year (Sheldrick et al., 2025).

Clinical thresholds or cut-off points in screening measures determine the classification of children and their level of risk for MH problems (Sheldrick et al., 2015). Determining appropriate clinical thresholds is key to early identification. If thresholds are too low, healthcare services risk being overwhelmed and parents unnecessarily concerned; too high, and at-risk children may be overlooked for intervention (Sheldrick et al., 2015). It is, therefore, important for researchers to determine appropriate thresholds and normative data based on the prevalence of child psychopathology and the population of interest. No normative data have

been established for the Preschool PSC. Whilst the PSC-17 has been the focus of a number of studies (McLean et al., 2025), more limited research on the psychometric properties of the Preschool PSC warrants further attention. Thus, this study includes an examination of both PSC measures that are appropriate for use with preschool populations.

4.2.4 Acceptability

Acceptability in terms of whether a measure is perceived as easy to comprehend, helpful or useful can affect implementation of screening (Brinley et al., 2024; Humphrey & Wigelsworth, 2016). Acceptability may increase uptake and implementation of screening programs and therefore, it is an important aspect to investigate if measures are to be completed as part of assessment and screening practices (Harrison et al., 2013). The acceptability of the Preschool PSC, as reported by any user, has not been explored. Investigating whether educators find the PSC measures acceptable is important if these measures are to be adopted in ECEC services.

4.2.5 The Current Study

No research has investigated the Preschool PSC with educators; little research has investigated incremental validity for either PSC measure; and no studies have investigated acceptability of the Preschool PSC. Thus, the overall aims were to examine validity and establish normative data for parent-reported Preschool PSC, examine the validity of educator-reported Preschool PSC and PSC-17, and examine acceptability of parent- and educator-reported Preschool PSC.

Two studies were conducted with two cohorts. The first study examined the validity and acceptability of parent-reported Preschool PSC, and the second study examined the validity

of educator-reported Preschool PSC and PSC-17, and incremental validity. Analyses were conducted using SPSS 29.0.1.0 (IBM SPSS, 2023).

4.3 Study 4a

Study 4a aimed to: (1) examine the psychometric properties of parent-reported Preschool PSC, in terms of internal consistency, test-retest reliability, concurrent and predictive validity, and to also examine scoring thresholds; (2) establish normative data for Preschool PSC in Australia; and (3) examine the acceptability of this measure.

4.3.1 Method

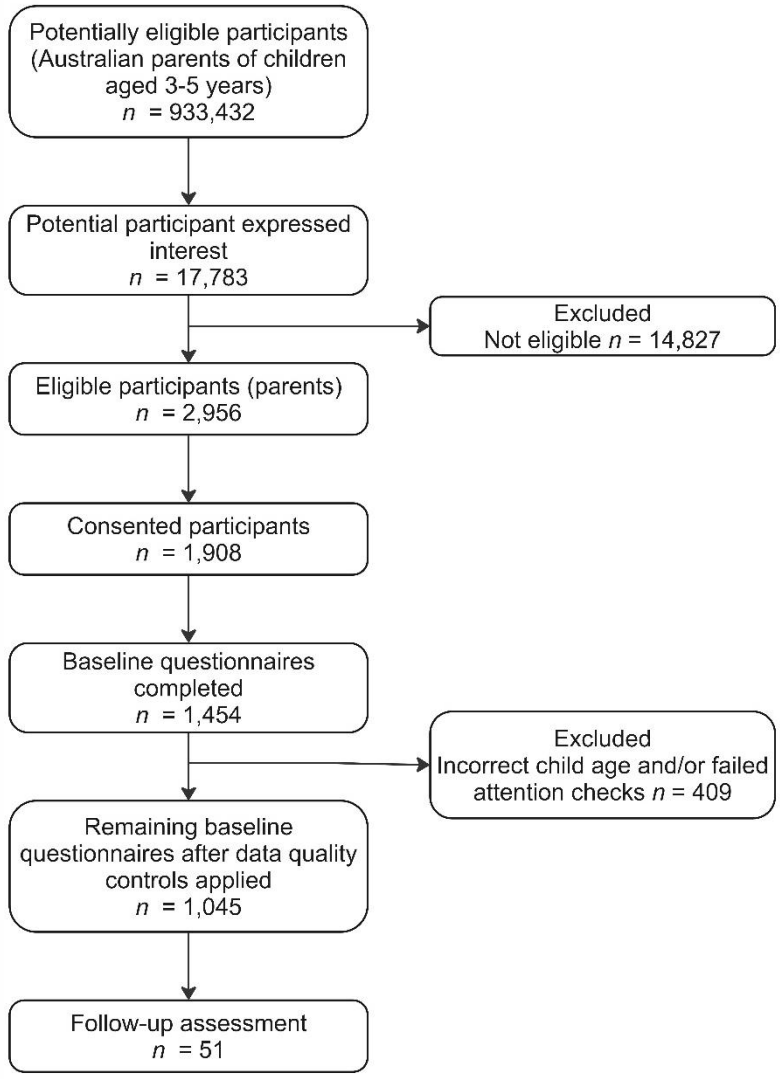
Procedures

Whilst there are no set conventions for sample sizes for normative data, this study aimed to recruit a large, representative sample from the Australian population using an online research panel. Participants were recruited through an Australian online research panel and sampled to represent families based on key socio-demographic characteristics from national census data (i.e., household income, marital status, and residential location) using segmented quotas. After sampling, parent demographic details, including household income, marital status, and residential location (metropolitan, regional/rural) were compared to census data to ensure key characteristics of the sample aligned with the broader population (Australian Bureau of Statistics, 2021a; 2022–23; 2023). Adult parents or caregivers of a child aged 3–5 years who lived in Australia and understood English completed baseline measures (N = 1,045), which were counterbalanced in order of completion. Participants who completed baseline questionnaires were invited to participate in a second assessment. To evaluate test-retest reliability, concurrent and predictive validity, a sub-sample of participants (n = 51) completed

a second set of questionnaires on average 35 days after baseline. Participant flow is presented in Figure 2. See Appendix A-7 for additional details about recruitment procedures.

Figure 2.

Flow of Participants through Study 4a



Participants

Mean child age was 3.87 years (*SD* = .78) and 46.89% were female. Child ethnicity was predominantly Caucasian, but diverse (see Appendix A-8 for details). Parents were aged 36.30 years on average (*SD* = 8.10); 66.6% of parents identified as female; and 84.1% of parents were married or in *de facto* relationships.

Measures

In addition to the Preschool PSC, parents' views on the acceptability of the Preschool PSC were assessed using four items adapted from a parent-report measure developed by Hawes et al. (2021). Responses were collected on a 5-point Likert response scale from *not at all/strongly disagree* (1) to *very/strongly agree* (5). Items index confidence in ratings ("Based on your knowledge of your child, how confident are you in the responses/ratings you have given today as part of the Checklist?"), the appropriateness ("How appropriate did you find the wording of the questions, i.e., were the questions appropriate for your child's age or developmental stage?" and "The Checklist asks important questions about my child's mental health") and clarity of item wording ("How clear did you find the wording of the questions, i.e., were they easy to understand?").

The CBCL for 1.5–5 years (Achenbach, 1999) was used as a criterion measure at the second assessment point to establish concurrent and predictive validity with Preschool PSC. The CBCL is a comprehensive, 99-item measure of child MH. Cases were classified as borderline, clinical, or at-risk including both borderline and clinical cases. The CBCL has strong psychometric properties and is the most common criterion measure in classification accuracy studies (Lavigne et al., 2016).

Analytic Plan

To examine the first aim, Cronbach's alpha for internal consistency and ICC for test-retest reliability were assessed. ICC estimates and their 95% confident intervals were calculated based on a single rating, absolute-agreement, 2-way mixed-effects model. To assess concurrent validity with the CBCL, Pearson correlations were calculated. Preschool PSC scores were classified into at-risk categories using the recommended scoring threshold. Receiver operator characteristic (ROC) analysis was used to evaluate concurrent validity based on categorical risk classifications and to examine recommended thresholds. Low cell counts meant assumptions were violated for chi-square tests. Fisher's Exact Test was not significant thus we were unable to analyse the predictive validity of the Preschool PSC compared to the CBCL clinical or at-risk cases as intended. Psychometric properties were evaluated according to the rubric developed by Youngstrom et al. (2017) and De Los Reyes and Langer (2018).

To address the second aim, normative data was determined by reporting scoring cut-offs for the 90th (borderline) and 95th percentile (at-risk) for overall age and gender (N = 1,045). To address the third aim of acceptability, frequency distributions were examined (n = 51).

Since the online survey required input for all items, there was no missing data.

4.3.2 Results

Reliability and Validity

Excellent internal consistency (α = .90) and adequate test-retest reliability was found for Preschool PSC total scores (ICC = .78, CI .65–.87).

Preschool PSC total scores strongly and positively correlated with CBCL total scores at the bivariate level, demonstrating concurrent validity, r = .86, p < .001, n = 51.

ROC analysis examined the association between Preschool PSC and CBCL at-risk case classifications. The area under curve (AUC) for Preschool PSC total scores was .89 with a standard error of .09. This was significant at $p < .001$ (CI .72–1.06).

ROC analysis identified an alternative clinical threshold of 19.5 for the Preschool PSC, which maximised sensitivity and specificity. For ease of scoring, the alternate score was rounded down and analysed at 19, which identified 10.0% ($n = 104$) of the sample as at-risk. Chi-square tests of independence indicated an adequate number of cell counts using the alternate threshold. An alternate clinical threshold using a criterion of CBCL clinical cases produced sensitivity of 83.3, specificity 95.6, PPV 71.4, NPV 97.7, and AUC .89 (standard error .09), $p < .001$.

Normative Data

The mean, standard deviation and ranges for total scores and subscales were calculated. Using the recommended scoring threshold, 47.4% of the sample ($n = 495$) were identified as at-risk. Table 10 presents scores for the highest-ranking cases by age and gender.

Acceptability

In response to questions of acceptability, most parents agreed or strongly agreed the Preschool PSC asked important questions about their child's MH (86.3%), found the Preschool PSC to be very or quite appropriate (86.3%) and very or quite clear (96.1%). Most parents were very or quite confident in the ratings they provided (96.1%).

Table 10.

Mean Scores and Banding for Preschool PSC by Age and Gender

		3–5 years		
		(*N* = 1,045)	Male (*n* = 554)	Female (*n* = 490)
Internalising	*M*	3.06	3.28	2.81
	SD	2.47	2.56	2.35
	Range	0–12	0–12	0–12
	Possible range	0–12		
	α	.777	.790	.757
	At-risk (top 5%)	8	8	7
	Borderline (top 10%)	7	7	6
Externalising	*M*	1.30	1.56	1.01
	SD	1.65	1.82	1.39
	Range	0–8	0–8	0–8
	Possible range	0–8		
	α	.776	.805	.705
	At-risk (top 5%)	4	5	4
	Borderline (top 10%)	4	4	3
Attention	*M*	1.95	2.21	1.67
	SD	1.72	1.81	1.57
	Range	0–6	0–6	0–6
	Possible range	0–6		
	α	.808	.825	.771
	At-risk (top 5%)	6	6	5
	Borderline (top 10%)	5	5	4
Parenting Challenges	*M*	3.24	3.55	2.88
	SD	2.91	3.02	2.75
	Range	0–12	0–12	0–12
	Possible range	0–12		
	α	.834	.844	.816
	At-risk (top 5%)	9	10	9
	Borderline (top 10%)	8	8	7
Total scores	*M*	9.22	10.21	8.10
	SD	6.81	7.28	6.07
	Range	0–36	0–36	0–36
	Possible range	0–36		
	α	.900	.911	.877
	At-risk (top 5%)	23	25	20
	Borderline (top 10%)	18.4	20	16.9
	Preschool PSC case rate (% positive)	47.4	53.1	40.8

Notes. Data not reported for one child identified as non-binary gender.

4.4 Study 4b

Having produced new validation evidence for the parent-reported Preschool PSC in Study 4a, Study 4b sought to extend measure utility by validating educator-reported Preschool PSC and PSC-17 measures and to examine the incremental value of multi-informant ratings by recruiting parents and educators.

The aims of Study 4b were to: (1) examine the psychometric properties of educator ratings of the Preschool PSC and PSC-17 in terms of internal consistency, test-retest reliability, inter-rater reliability, predictive validity, concurrent validity and convergence with diagnostic clinical interviews; (2) examine the incremental validity of parent- and educator-report Preschool PSC and PSC-17 scores using criterion measures including the parent-report CBCL and clinician-rated CGAS; and (3) examine acceptability of the measures amongst educators.

4.4.1 Method

Procedures

An *a priori* statistical power calculation to estimate an appropriate sample for analysing measurement stability and validity was performed using G*Power 3.1.9.7 (Faul et al., 2009). With alpha of .05, power .80, and an effect size of 0.12, the calculated sample size was 93. Participants were recruited via ECECs and social media advertisements. Adult parents or caregivers of a child aged 3–5 years, who lived in Australia, and understood English were asked for permission to contact their child's educator, as part of the consent process. Eligible educators were adult educators of children aged 3–5 years, who worked in ECEC, had internet access, understood English and knew the child well enough to complete questionnaires about their MH. The final paired sample of parents and educators who completed baseline questionnaires was 94.

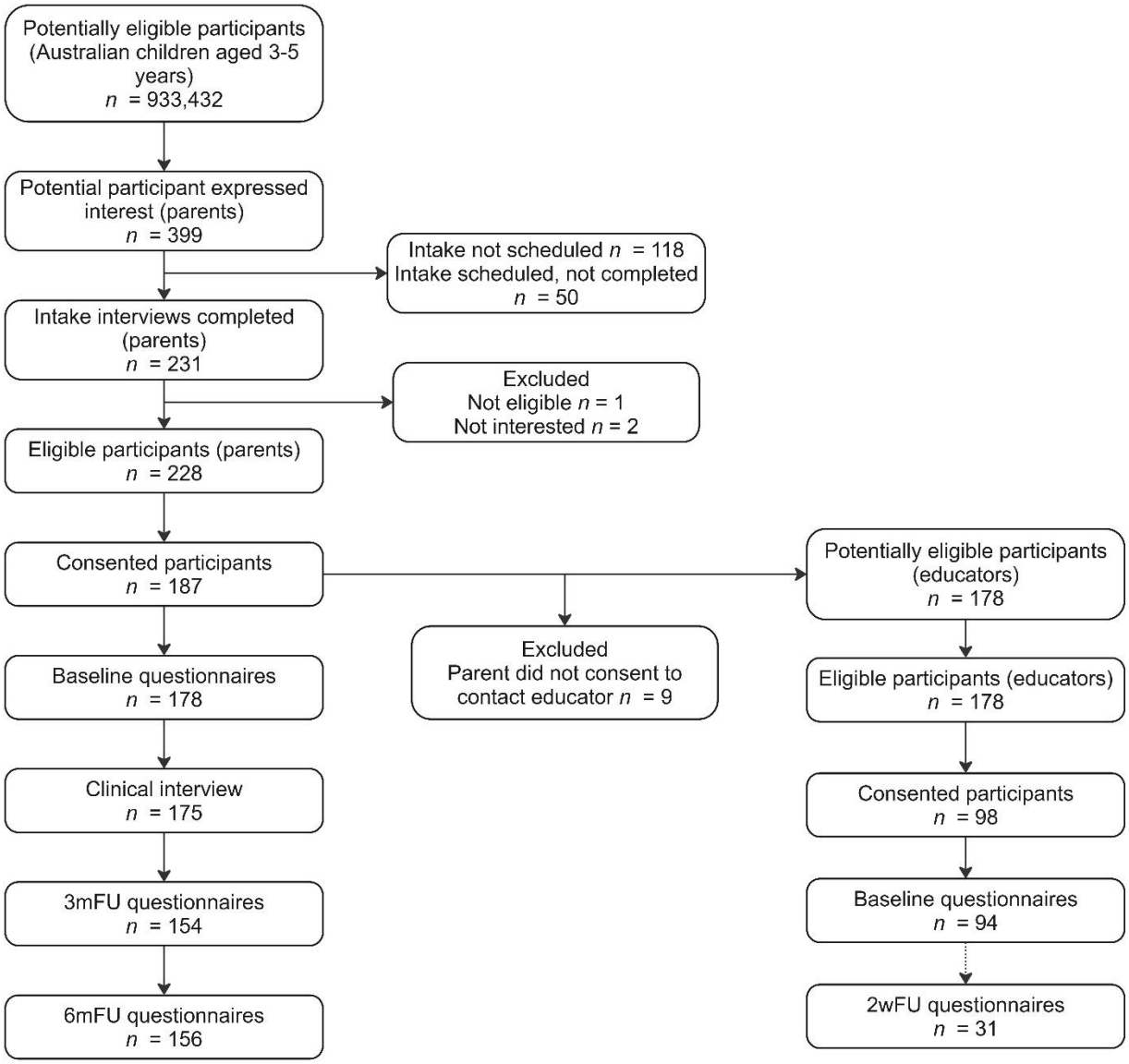
Participants completed measures which were counterbalanced in order of completion at each time point. Parents completed baseline, 3-month (3mFU) and 6-month follow-up (6mFU) questionnaires. Within two weeks of baseline, parents completed one clinical interview (2wFU), and educators completed baseline questionnaires. Within two weeks of educator baseline, a sub-sample of educators (n = 31) completed a second set of questionnaires to test test-retest reliability. Educators who completed baseline questionnaires were invited to participate in the second assessment until the target recruitment for the subsample was reached. Participant flow is presented in Figure 3.

Participants

Children were aged 3.65 years on average (SD = 0.67) and 38.3% female. Child ethnicity was predominantly Caucasian (74.5%), followed by 16% other ethnicities. Educators were predominantly female (97.9%) with a mean of 13.13 years ECEC work experience (SD = 8.89). Most educators indicated that they were *very* or *quite* familiar with the child (89.4%, n = 84) and had known the child between 2–54 months (M = 13 months). See Appendix A-9 for additional participant details.

Figure 3.

Flow of Participants through Study 4b



Measures

Educators' views on the acceptability of measures were assessed at baseline using the same measure as reported in Study 4a.

Parents completed the CBCL at baseline, 3mFU and 6mFU. The CBCL at baseline and 3mFU was used in this study as a criterion measure to investigate convergent and predictive validity, and as an outcome measure for incremental validity.

Clinicians, who were blind to parents and educators' questionnaire responses, conducted clinical interviews with parents using the Diagnostic Interview Schedule for Children, Adolescents and Parents (DISCAP-V; Tissue et al., 2022) to assess child psychiatric disorder and the Children's Global Assessment Scale (CGAS; Shaffer et al., 1983) to assess global functioning. The DISCAP-V is a semi-structured interview guide corresponding to common childhood behavioural and affective diagnostic categories. Clinicians rated the severity of the diagnostic presentation (0–6). Those with a severity rating 4 and above were considered to meet criteria for the specified disorder. The CGAS is an outcome measure of a child's general functioning rated on a continuum (0–100); higher scores indicate higher levels of health and superior functioning, while lower scores indicate illness and gross impairment. Parents were blind to clinician ratings.

Clinicians were clinical psychology interns at a university research clinic, registered general and clinical psychologists, all under the supervision of senior clinical psychologists. Inter-rater reliability for a subset of clinical interviews ($n = 24$) was assessed using a secondary team of clinicians, who were blind to the primary clinician's ratings. Diagnostic ratings correlated highly ($r = .79$, $p < .001$). CGAS scores moderately correlated ($r = .57$, $p = .004$). There were no significant differences in ratings.

Educators' familiarity with the child was assessed using five items such as how frequently they care for the child, how long they have known the child, level of familiarity and

strength of relationship with the child. Responses were collected on a 5-point Likert response scale from *not at all/strongly disagree* (1) to *very/strongly agree* (5).

Parent-report Preschool PSC was completed at baseline, 3mFU and 6mFU. One item differed for educator-report Preschool PSC: "is it hard to take the child out in public?" with an additional response option, *Not applicable*. Scoring for this response option was coded as 0. Items were summed to produce a total score. Educators completed this measure at baseline and 2wFU.

The PSC-17 was the same for parents and educators. Parents completed this measure at baseline, 3mFU and 6mFU and educators at baseline and 2wFU.

Analytic Plan

The following sub-samples of participants were used to analyse the data: baseline educator (n = 94), educator 2wFU (n = 31), baseline parent (n = 178), parent 2wFU (n = 175), and parent 3mFU (n = 154). To examine the first aim, Cronbach's alpha was calculated to assess reliability using baseline educator assessment, ICCs were assessed for test-retest (baseline educator and 2wFU) and inter-rater reliability, and Pearson's correlations for convergent validity with CBCL total scores (baseline parent and educator). ICC estimates and 95% confident intervals were calculated based on a single rating, absolute-agreement, 2-way mixed-effects model. To assess inter-rater agreement, Cohen's κ was calculated using case classifications (baseline parent and educator). The two PSC measures were classified using recommended thresholds. To test concurrent validity with the DISCAP, a series of binary logistic regressions were conducted (baseline educator and parent 2wFU). To assess predictive validity, first cell counts for chi-square tests and Fisher's Exact Test were checked to be valid. Baseline educator Preschool PSC and PSC-17 ratings were then compared to parent-rated

CBCL clinical cases at 3mFU. Psychometric properties were evaluated according to the same rubric as utilised in Study 4a.

To address the second aim, a series of hierarchical multiple regressions were conducted to assess incremental validity (baseline educator, baseline parent and parent 2wFU). To address the third aim of acceptability, frequency distributions were examined for educators at baseline.

Since the online survey required input for all items, there was no missing data for baseline measures for parents and educators, and educator 2wFU. Listwise deletion was used for two cases where parents did not attend clinical interviews. Pairwise deletion was used for parent-reported data missing at 3mFU (n = 13).

4.4.2 Results

Reliability and Validity

Internal consistency was excellent for baseline educator-reported Preschool PSC (α = .91) and good for PSC-17 total scores (α = .88). Test-retest reliability for educator-reported Preschool PSC was excellent (ICC = .91; CI .82–.96, p < .001) and good for the PSC-17 (ICC = .80; CI .66–.89, p < .001) for baseline and 2wFU assessments. Inter-rater reliability between baseline parent and baseline educator ratings was small for the Preschool PSC (ICC = .33; CI .14–.49, p < .001) and PSC-17 (ICC = .34, CI .15–.50, p < .001).

Convergent validity was weak and insignificant for baseline educator-reported Preschool PSC and baseline parent-reported CBCL total scores, r = .19, p = .068, and baseline educator-reported PSC-17 total scores, r = .19, p = .067. Cohen's κ was run to determine if there was agreement between at-risk baseline educator-reported Preschool PSC and baseline parent-reported CBCL at-risk cases, κ = .15, CI -.05–.36, p = .123. When at-risk baseline educator-reported Preschool PSC cases were compared to baseline parent-reported CBCL clinical cases, there was significant, though low, agreement, κ = .19, CI -.01–.40, p = .040. At-

risk baseline educator-reported PSC-17 cases compared to baseline parent-reported CBCL at-risk cases found κ = .11, CI -.11–.33, *p* = .275, and to baseline parent-reported CBCL clinical cases κ = .09, CI -.14–.33, *p* = .360.

To test concurrent validity, a series of univariate binary logistic regressions were performed with baseline educator-reported Preschool PSC and PSC-17 total scores, and at-risk classifications as the predictors and any type of DISCAP diagnosis as the outcome (*n* = 92). None of the models were significant as presented in Table 11. Neither PSC significantly contributed towards diagnosis.

Table 11.

Tests of Convergence for Baseline Educator-report PSC Measures with DISCAP Diagnoses

Predictor variables Reference: DISCAP diagnosis *n* = 92	***χ2*, Significance**	**Nagelkerke R^2**	**Wald**
PPSC total scores	1.77, *p* =.183	0.03	1.84, *p* = .175
PPSC at-risk classifications	1.20, *p* = .273	0.02	1.22, *p* = .269
PSC-17 total scores	1.19, *p* = .276	0.02	1.21, *p* = .271
PSC-17 at-risk classifications	1.26, *p* = .268	0.02	1.31, *p* = .253

Notes. PPSC = Preschool PSC. Coding for case classifications: 0 = not at-risk, 1 = at-risk. Coding for DISCAP diagnosis: 0 = no diagnosis, 1= one or more diagnoses.

Predictive validity

The predictive validity of the educator-reported measures was tested against the CBCL using chi-square tests. Compared to parent-reported CBCL clinical cases at 3mFU, baseline educator-reported Preschool PSC scores produced sensitivity 24.0%, specificity 89.3%, PPV 50.0 and NPV 72.46; and PSC-17 scores produced sensitivity 36.4%, specificity 88.6%, PPV 33.33 and NPV 89.86. Classification accuracy as indicated by AUC was moderate for the Preschool PSC (AUC = .65, CI .51–.79, p = .037) and PSC-17 (AUC = .63, CI .49–.77, p = .061).

Incremental validity

To test incremental validity, hierarchical multiple regressions were conducted, with three blocks of variables. First, child age and gender were entered as predictors with clinician-rated CGAS as the outcome variable; neither were significant, $F(2) = 1.15$, $p = .321$, $R^2 = .02$. In the second model, baseline parent-reported Preschool PSC total scores were added and showed significant improvement on the first model $F(3) = 14.19$, $p < .001$, $R^2 = .33$. In the third model, baseline educator-reported Preschool PSC total score was added and was significant $F(4) = 11.77$, $p < .001$. It explained an additional 2.5% of variance; however, this was not a significant change, indicating no increase in incremental validity when educator reports were included.

When the order of entry was reversed and baseline educator-report Preschool PSC scores were entered in the second model and baseline parent-report scores in the third model, both models were found to significantly improve prediction of clinician-rated CGAS scores after controlling for age and gender, indicating an increase in incremental validity when parent reports were included. Hierarchical regressions were also performed, with parent-reported CBCL as the outcome variable. Overall, models using the Preschool PSC explained 35.1% in

variance of CGAS scores and 50% of CBCL total scores. Table 12 presents a summary of results.

Analysis was repeated with the PSC-17 and clinician-rated CGAS as the outcome variable. After controlling for age and gender, baseline parent-reported PSC-17 was a significant predictor, $F(3) = 4.75$, $p = .004$, and explained 13.9% of variance. The third model was also significant, $F(4) = 4.58$, $p = .002$ and explained an additional 3.5% variance, however this was not a significant change in incremental validity when educator reports were included.

When the order of entry was reversed, both models were found to significantly improve prediction of clinician-rated CGAS scores. After controlling for age and gender, baseline educator-report PSC-17 explained 11.3% of variance, $F(3) = 3.73$, $p = .014$. The third model was also significant, $F(4) = 4.58$, $p = .002$ and explained an additional 6.1% of variance, $p = .013$ indicating a significant increase in incremental validity when baseline parent reports were included. Overall, models using the PSC-17 explained 17.4% in variance of CGAS scores and 41% of CBCL total scores. Table 13 presents a summary of results.

Acceptability

In response to questions of acceptability, educators found the PSC measures highly acceptable. Educators agreed or strongly agreed the measures asked important questions about the child's MH (Preschool PSC 78.7%, $n = 74$; PSC-17 87.2%, $n = 82$), found the measures to be very or quite appropriate (Preschool PSC 84%; PSC-17 86.1%) and very or quite clear (Preschool PSC 94.7%; PSC-17 93.6%). Educators were very or quite confident in the ratings they provided (Preschool PSC 90.4%; PSC-17 93.6%).

Paired sample t-tests showed there was no significant difference between educator ratings of acceptability between the Preschool PSC and PSC-17 in terms of appropriateness or clarity. In terms of whether the measures asked important questions, educators rated the PSC-

17 significantly higher on average (M = 3.15, SD .69) than the Preschool PSC (M = 2.98, SD .73), $t(93) = -3.05$, $p = .003$.

Table 12.

Summary of Hierarchical Regression Models for Educator-Report PPSC and Parent-Report PPSC Total Scores as Predictors

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CGAS									
Block 1:	.03	.03	1.15	2	.321				
Age						-.21	.14	-.16	.134
Gender						-.72	2.23	-.03	.748
Block 2:	.33	.30**	14.19	3	< .001**				
Age						-.17	.12	-.13	.144
Gender						-1.77	1.87	-.08	.347
Parent-reported PPSC						-1.12	.18	-.55	< .001**
Block 3:	.35	.03	11.77	4	< .001**				
Age						-.20	.11	-.16	.083
Gender						-1.97	1.85	-.09	.289
Parent-reported PPSC						-.99	.19	-.49	< .001**
Educator-reported PPSC						-.26	.14	-.17	.070

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CGAS									
Block 1:	.03	.03	1.15	2	.321				
Age						-.21	.14	-.16	.134
Gender						-.72	2.23	-.03	.748
Block 2:	.15	.12**	5.06	3	.003*				
Age						-.26	.13	-.20	.045*
Gender						-1.39	2.10	-.07	.512
Educator-reported PPSC						-.54	.15	-.35	< .001**
Block 3:	.35	.20**	11.77	4	< .001**				
Age						-.20	.11	-.16	.083
Gender						-1.97	1.85	-.09	.289
Educator-reported PPSC						-.26	.14	-.17	.070
Parent-reported PPSC						-.99	.19	-.49	< .001**
CBCL									
Block 1:	.01	.01	.61	2	.546				
Age						.10	.28	.04	.730
Gender						4.92	4.52	.12	.279
Block 2:	.49	.48**	29.15	3	< .001**				
Age						-.002	.20	-.001	.994
Gender						7.62	3.27	.18	.022*
Parent-reported PPSC						2.88	.31	.70	< .001**
Block 3:	.50	.003	21.89	4	< .001**				
Age						-.02	.20	-.01	.904
Gender						7.47	3.28	.17	.025*
Parent-reported PPSC						2.97	.34	.72	< .001**
Educator-reported PPSC						-.19	.25	-.06	.460

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CBCL									
Block 1:	.01	.01	.61	2	.546				
Age						.10	.28	.04	.730
Gender						4.92	4.52	.12	.279
Block 2:	.05	.04*	1.73	3	.167				
Age						.16	.28	.06	.555
Gender						5.71	4.47	.13	.204
Educator-reported PPSC						.64	.32	.21	.051*
Block 3:	.50	.44**	21.89	4	< .001**				
Age						-.02	.20	-.01	.904
Gender						7.47	3.28	.17	.025*
Educator-reported PPSC						-.19	.25	-.06	.460
Parent-reported PPSC						2.97	.34	.72	< .001**

Notes. Table presents a summary of hierarchical regression models for baseline educator-report Preschool PSC and baseline parent-report Preschool PSC total scores as predictor variables with CGAS and CBCL total scores as outcome variables ($n = 92$). PPSC = Preschool PSC. *Significant at $p < .05$, ** $p < .01$

Table 13.

Summary of Hierarchical Regression Models for Educator-Report PSC-17 and Parent-Report PSC-17 Total Scores as Predictors

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CGAS									
Block 1:	.03	.03	1.15	2	.321				
Age						-.21	.14	-.16	.134
Gender						-.72	2.23	-.03	.748
Block 2:	.14	.11**	4.75	3	.004*				
Age						-.15	.13	-.12	.254
Gender						-1.97	2.14	-.09	.358
Parent-reported PSC-17						-.85	.25	-.35	< .001**
Block 3:	.17	.03	4.58	4	.002*				
Age						-.19	.13	-.15	.146
Gender						-2.41	2.12	-.12	.258
Parent-reported PSC-17						-.67	.26	-.27	.013*
Educator-reported PSC-17						-.36	.19	-.20	.060

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CGAS									
Block 1:	.03	.03	1.15	2	.321				
Age						-.21	.14	-.16	.134
Gender						-.72	2.23	-.03	.748
Block 2:	.11	.09*	3.73	3	.014*				
Age						-.25	.13	-.19	.062
Gender						-1.77	2.17	-.08	.416
Educator-reported PSC-17						-.53	.18	-.30	.004*
Block 3:	.17	.06*	4.58	4	.002**				
Age						-.19	.13	-.15	.146
Gender						-2.41	2.12	-.12	.258
Educator-reported PSC-17						-.36	.19	-.20	.060
Parent-reported PSC-17						-.67	.26	-.27	.013*
CBCL									
Block 1:	.01	.01	.61	2	.546				
Age						.10	.28	.04	.730
Gender						4.91	4.52	.11	.279
Block 2:	.41	.39**	20.47	3	< .001**				
Age						-.12	.22	-.05	.584
Gender						9.66	3.58	.23	.008*
Parent-reported PSC-17						3.20	.42	.64	< .001**
Block 3:	.41	.00	15.19	4	< .001**				
Age						-.13	.22	-.05	.569
Gender						9.58	3.62	.22	.010*
Parent-reported PSC-17						3.24	.45	.65	< .001**
Educator-reported PSC-17						-.06	.32	-.02	.846

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CBCL									
Block 1:	.01	.01	.61	2	.546				
Age						.10	.28	.04	.730
Gender						4.91	4.52	.11	.279
Block 2:	.06	.05*	1.88	3	.139				
Age						.16	.27	.06	.562
Gender						6.47	4.50	.15	.154
Educator-reported PSC-17						.78	.37	.22	.039*
Block 3:	.41	.35**	15.19	4	< .001**				
Age						-.13	.22	-.05	.569
Gender						9.58	3.62	.22	.010*
Educator-reported PSC-17						-.06	.32	-.02	.846
Parent-reported PSC-17						3.24	.45	.65	< .001**

Notes. Table presents a summary of hierarchical regression models for baseline educator-report PSC-17 and baseline parent-report PSC-17 total scores as predictor variables with CGAS and CBCL total scores as outcome variables ($n = 92$).
*Significant at $p < .05$, ** $p < .01$

4.4.3 General Discussion

With an aim of examining the psychometric properties of parent-reported Preschool PSC, the first study contributed new reliability and validity evidence for the measure, finding excellent internal consistency, adequate test-retest reliability and strong concurrent validity with the CBCL total scores and case classifications. Given the high proportion of children identified as at risk in this sample using the original threshold, alongside consideration of new normative data addressing the second aim, an alternative threshold was identified as 19, which is substantially higher than the original cut-off of 9. High sensitivity and specificity using alternative thresholds surpassed benchmark standards for developmental screening measures (Aylward, 1997; Lavigne et al., 2016).

This study was the first to explore acceptability of the parent-rated Preschool PSC—the study's third aim. Acceptability of the Preschool PSC by parents was very high. More than 85% of parents rated the Preschool PSC as appropriate on all indices. Our results align with previous findings with older children across health and education settings, in which parents perceived universal MH screening as appropriate and helpful (Brinley et al., 2024).

The second study complemented the first by examining the psychometric properties of educator-reported Preschool PSC and PSC-17. Very high internal consistency and test-retest reliability was found for both measures. Unsurprisingly, inter-rater reliability between parent and educator ratings was small, but significant for both measures, which corresponds with previous findings in which there is low correspondence between informants when they observe child behaviour in different settings (De Los Reyes et al., 2015). Specifically for the PSC-17, the correspondence between parent and educator ratings, whilst small (ICC = .34, CI. 15–.50, $p < .001$), was higher than found in a previous study on the measure, which reported a bivariate correlation of $r = .26$, $p < .001$ (Moore, Dowdy, Fleury, et al., 2022). This was the first study to test inter-rater reliabilities for parent and educator reports on the Preschool PSC, so there are

no comparisons to make in the literature. This study's results are comparable to other studies of agreement between parent and teacher reporters on screening measures of anxiety (r .21 - .32, $p < .01$) (Reardon et al., 2018), although cross-informant agreement does vary widely across measures (McLean et al., 2024). For analyses in which the outcome variable was parent-reported CBCL or clinician-rated diagnosis (concurrent, convergent, predictive and incremental validity), educator ratings did not perform well overall, although educator classification accuracy was moderate, as indicated by AUC analysis and educator ratings for both PSC measures had high specificity. These results may reflect informant discrepancies, possible bias of the parent-report criterion measure or criterion contamination.

Addressing the second aim of incremental validity, results indicated that incremental validity of parent report in addition to educator report was demonstrated for both PSC measures when the outcome was parent-reported CBCL scores. Educator reports did not add significant variance for the parent-reported outcome. Educator-reported PSC measures significantly improved the prediction of clinician-rated functioning scores, only when educator report was entered before parent report. In contexts where educators are the first gate, such as an ECEC setting, parents add significantly more variance to the model, which contrasts with previous research that found educators added significantly more information on top of parent ratings for the PSC-17 (Moore, Dowdy, Fleury, et al., 2022). Interestingly, the PSC-17 did not explain as much variance in these models, suggesting the Preschool PSC's items assessing parenting challenges may more closely align with impairment and functioning. These findings signal the importance of gathering parent ratings on child MH in addition to educator reports. Despite challenges facing time-poor caregivers, this research highlights the value of multi-informant assessment in young children.

To address the final aim, educators rated both measures as highly acceptable. Interestingly, educators rated the PSC-17 more highly than the Preschool PSC in terms of

whether it asked important questions; a difference not reflected in parent ratings. Despite concerns that some PSC-17 items are not well-suited to this age group, results from both studies indicate the PSC-17 was perceived by parents and educators as being age-appropriate and highly acceptable. These findings align with previous research, which found parents and educators find screening programs acceptable, and parents of preschool-age children rate the PSC-17 as an acceptable screening measure (Brinley et al., 2024; McLean et al., 2024). Findings from the present study that both PSC measures were acceptable to parents and educators support the use of these measures with caregivers of preschool-age children.

Study strengths include the large national sample complemented by a paired parent and educator sample. This was the first to validate an educator-report version of the Preschool PSC and utilise best-available, gold-standard criteria such as diagnostic interview. However, the large sample could not be used for all analyses and sub-samples were small. Future research may include random sampling to ensure the generalisability of results, and inviting all educators to complete follow-up questionnaires. Community recruitment meant few clinical cases were identified in the first study, which affected predictive validity analysis. Further, parent-reported CBCL may have been a source of criterion contamination; results may have biased parent-rated predictors. This study did not include either a comprehensive educator-report questionnaire nor a clinical assessment including interviews with educators, which may have biased results towards parent reports and not sufficiently accounted for MH symptoms displayed outside the home context. Future research may mitigate this by employing comprehensive educator-report measures or having clinicians interview educators to inform diagnostic ratings. Future studies may seek to recruit larger educator samples and replicate this study's findings.

In terms of research implications, our findings support the Preschool PSC as a valid, reliable and acceptable measure of child MH for preschool-age children and add weight to the

suitability of using the PSC-17 in the preschool-age range with educator report. As free, brief measures, the Preschool PSC and PSC-17 are important offerings in longitudinal research, with clear benefits enabling children to be tracked longitudinally using a single measure.

The first study found that a sizable proportion of children were at risk using the existing Preschool PSC clinical threshold of 9 (Sheldrick, Henson, Merchant, et al., 2012) suggesting this threshold is too low and may overestimate children at risk of MH problems. The existing clinical threshold identified 47.4% of preschool-age children as at-risk for MH problems, whereas the most recent National Survey of Mental Health and Wellbeing in Australia identified 14% of children aged 4–17 years as scoring in the clinical range of the CBCL (Sawyer et al., 2001). A similar survey in Western Australia which compared prevalence in older children (12–16 years) to younger children (4–11 years) reported 16% prevalence in younger children (Garton et al., 1998). The alternative Preschool PSC threshold of 19 identified 10.0% of the sample as at-risk. Thus, this revised clinical threshold appears to be more appropriate to use in young Australian children and should be adopted in future research and clinical practice using the Preschool PSC. Another important implication for determining appropriate thresholds is the level of available healthcare resources that can manage the number of children identified by screening measures as requiring early intervention (Sheldrick & Garfinkel, 2017). Without sufficient resources to manage identified at-risk children, iatrogenic effects such as family burden and over-stretched healthcare systems may arise as a result of widespread screening programs (Sheldrick & Garfinkel, 2017).

Current universal screening and assessment practices in the preschool-age range focus predominantly on school readiness and academic competency (Cook et al., 2010; Sim, Thompson, Marryat, Ramparsad, et al., 2019); however, there is a clear need for screening to be extended to child MH, given the high rates of children at risk of problems. The clinical relevance and potential application of the overall findings for both PSC measures indicate

caregivers may be willing to adopt screening measures in their educational and care practice. ECEC services, which are less encumbered by issues of access and stigma (Desta et al., 2017), may be an ideal setting for early identification and can make use of educators as crucial early identifiers of child MH problems. Providing educators with validated tools as part of wider early intervention programs may help identify children in need of support. However, additional training programs are also needed to help educators communicate with parents about child wellbeing and supporting further help seeking (Paton et al., 2022). Training educators to support children's MH and to support appropriate parental help-seeking for children is essential, given parents frequently seek help for their child's MH problems from a preschool staff member in the first instance (Pavuluri et al., 1996). Information about the child from the home and education setting not only contributes to an ideal model of assessment but also may help parents recognise the need for further support for their child's wellbeing, although further research is required.

This research has expanded the evidence base for the validity and acceptability of parent- and educator-reported Preschool PSC and PSC-17 as utilised with young children. It presents new evidence using multi-informant reports, justifying the collection of data from educators. By validating these measures and establishing normative data, this research has increased the clinical utility of the PSC measures, enabling the early identification of young children at risk of MH problems.

CHAPTER 5: Longitudinal Stability and Validity of Parent-reported Mental Health Measures in a Preschool Community Population

The contents of this chapter have been submitted for publication in *Parenting: Science and Practice*.

5.1 Chapter Synopsis

Parent-report screening measures, which accurately capture socio-emotional or behavioural (SEB) symptoms in preschool-age children, are needed to identify children requiring support. This study aimed to investigate the assessment stability of SEB symptoms in a preschool community population, examine the validity of two screening measures, and examine if concurrent validity differed by age. Parents (N = 178) of children aged 3–5 years (41% female) from the community, completed the PSC-17, Preschool PSC and CBCL at baseline, 3-month and 6-month follow-up. At 2-week follow-up, clinicians interviewed parents using the Diagnostic Interview Schedule for Children, Adolescents and Parents (DISCAP-V), and CGAS. Preschool PSC total scores (ICC = .64–.83), PSC-17 total scores (ICC = .59–.67), externalising (ICC = .54–.55) and attention subscales (ICC = .75–.79) had significant measurement stability across time. The Preschool PSC showed increasing stability as children aged whilst the PSC-17 showed decreasing stability across time points. Moderate–strong associations of functioning scores with the Preschool PSC (r = -.59) and PSC-17 (r = -.49) supported concurrent validity. Both PSC measures had moderate–strong correlations with clinical diagnoses, with the exception of internalising disorders. Predictive validity of PSC measures with DISCAP-V diagnoses resulted in mixed findings. Concurrent validity with clinical diagnoses did not differ by age. This study adds to the evidence base that SEB symptoms are stable during the preschool period and parent-report screening using the PSC

measures is valid for preschool children. Findings that the measures are associated with clinical diagnoses further justifies their use as a low-cost first step in early intervention models.

5.2 Introduction

Normative development in young children can include challenging social-emotional behaviours. Whilst many of these behaviours are developmentally appropriate and will diminish as the child matures, the persistence or worsening severity of such behaviours can be indicative of more serious SEB concerns. Understanding the typical trajectory of symptoms in early childhood is important for identifying children with elevated SEB symptoms. Preschool children aged 3–5 years with elevated SEB symptoms are more likely to have health, language and development problems at age six, require additional educational supports throughout school, and have later adult MH diagnoses (Barbarin, 2007; Mulraney et al., 2021; Sim, Thompson, Marryat, Law, et al., 2019). Therefore, it is important to identify SEB difficulties in young children so that they can receive further assessment and support. In order to identify and refer children for further support, there is a need for parent-report measures which accurately capture SEB symptoms.

Whilst the measurement of SEB wellbeing can be challenging during the preschool period due to developmentally appropriate fluctuation in symptoms, SEB symptoms indicative of longitudinal risk can be detected through parent reports in children as young as 18–24 months (Bayer et al., 2012; Tremblay et al., 2004). Elevated SEB symptoms can be indicated by scores which exceed clinical thresholds on validated measures and can identify children who are at risk of further SEB problems and may require support. Elevated symptoms in childhood, more so than diagnosis, are associated with adult MH disorder highlighting the importance of identifying at-risk children (Mulraney et al., 2021).

Parent-report screening measures are a time- and cost-efficient method of identifying children who have elevated symptoms and may be at risk of SEB problems. Two broadband child MH measures in particular have been highlighted as having excellent psychometric accuracy (Becker-Haimes et al., 2020), the PSC and the SDQ. Both measures have been used to examine the stability of symptoms in preschool-age populations (Klein et al., 2015; Sheldrick et al., 2025). The SDQ, however, involves costs for digital use, which in the context of universal screening or large population studies can present a significant financial barrier. Alternatively, the PSC is free to use in both paper-based and digital formats.

The PSC-17 (Gardner et al., 1999) is a 17-item version of the PSC that produces a total score and three subscale scores based on internalising problems, such as inner, emotional distress; externalising problems such as conflict with others; and attention problems such as difficulties concentrating. The Preschool PSC is a related measure, adapted from the PSC-17, which targets the same three domains, with age-appropriate wording changes, plus an additional "Parenting challenges" domain that assesses aspects of parenting such as difficulties caring for the child. It comprises 18 items and produces a total score. Whilst both PSC measures are widely used in research and clinical care settings, few longitudinal studies have investigated child SEB symptoms in community preschool populations. Conducting this work is important for establishing the stability of screening measure assessments in young, community populations.

The Preschool PSC and PSC-17 have been validated for use with children aged 3–5 years (McLean et al., 2025a; McLean et al., 2025b), however, the accuracy of these two measures with preschool children in comparison to the best-available, gold standard in psychiatric diagnosis, that is, diagnostic clinical interviews, has yet to be verified. Validation against clinical interviews is important to address issues of common method variance, which

are possible limitations of previous studies using comprehensive questionnaires such as the CBCL (Achenbach, 1999).

To address these issues, the current study asks what is the stability of child MH symptoms in a preschool community sample as measured by screening measures; do the PSC measures correspond with clinical assessments; and do PSC measures differ in level of association with criterion measures at different ages? Thus, the aims of the study were firstly, to investigate the stability of parent-reported SEB symptoms in a community sample of preschool children using the PSC-17, Preschool PSC and CBCL. Second, the study aimed to investigate the association (concurrent and predictive) of parent-rated Preschool PSC and PSC-17 scores with clinical diagnoses and ratings of child functioning. Finally, the study aimed to investigate whether the Preschool PSC and PSC-17 differ in level of association in terms of strength of relationship with the CBCL and clinical diagnoses at different ages.

5.3 Method

Procedures

Participants were recruited as part of a larger study investigating parent and educator reports of child MH for preschool-age children. The University of Sydney Ethics Committee approved the study (2023/704). Study information was distributed via preschools and social media advertisements. All parents, with and without concerns about their child's MH, were encouraged to participate. Eligible participants were adult parents or caregivers (N = 178) of a child aged 3–5 years, who lived in Australia, and could complete measures in English. Participants provided informed consent prior to participation.

In one sitting online, parents completed questionnaires. Measures were counterbalanced in order of completion at each time point. Parents completed the same measures at baseline, 3-month (3mFU) and 6-month follow-up (6mFU). Questionnaires took

approximately 10–15 minutes to complete at each time point. Within two weeks of baseline, parents completed one clinical interview via phone or telehealth (2wFU) with a psychologist. Parents who completed baseline questionnaires were invited to complete all assessment time points regardless of whether they had completed previous time points. Parents who completed all time points and passed attention checks were sent a gift voucher at completion of the study (n = 133). Figure 4 presents participant numbers at each stage of the study.

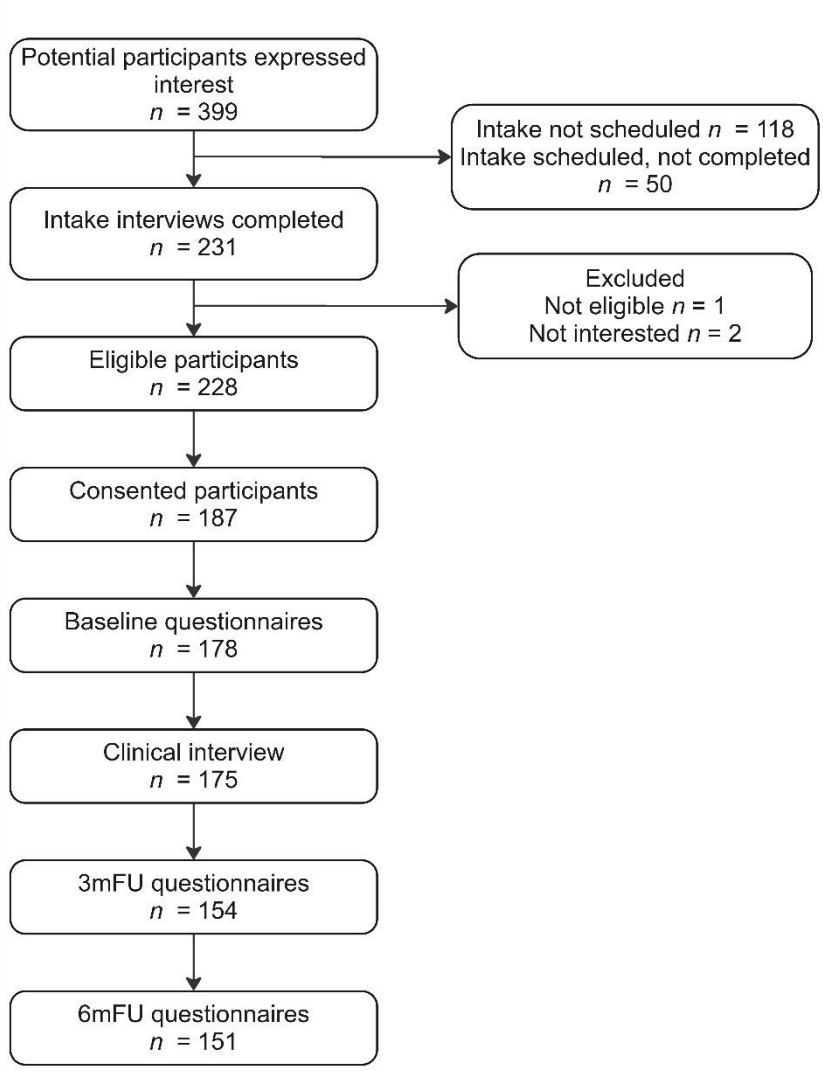
Children were aged 48.05 months on average (SD = 7.62) and 41% female. Child ethnicity was predominantly Caucasian (68.5%), followed by multiple ethnicities (16.3%) and South-East Asian (4.5%). Parents were 37.13 years old (SD = 4.61), 95.5% female and educated (84.3% held a university-level qualification).

Measures

The PSC-17 has been recently validated for children aged three, and reported to have good internal consistency, excellent concurrent validity, good predictive validity for preschool-age children and high acceptability amongst parents (McLean et al., 2025). The Preschool PSC also has strong internal consistency, good test-retest reliability, strong concurrent validity, and moderate to good predictive validity (McLean et al., 2025b; Rocha & Nunes, 2021; Sheldrick, Henson, Merchant, et al., 2012; Sheldrick et al., 2025). Amongst parents it also has reported high acceptability in terms of being an appropriate measure to use (McLean et al., 2025b). Both PSCs were completed at baseline, 3mFU and 6mFU.

Figure 4.

Flow of Participants through Study 5



The CBCL for 1.5–5 years (Achenbach, 1999) is a comprehensive, 99-item measure of child MH and was used as a criterion measure in this study. It produces total problems, syndrome and subscale scores. This study utilised the aggressive and attention syndromes, and internalising subscale. It has strong psychometric properties and was completed by parents at baseline, 3mFU and 6mFU. The CBCL was used in this study to address aims one and three; as a comparison measure to assess the longitudinal stability of measures, and to examine the strength of relationship with the PSCs at different ages.

Clinical interviews were utilised to address aims two and three. Psychologists conducted interviews with parents within two weeks of baseline questionnaires, using the Diagnostic Interview Schedule for Children, Adolescents and Parents (DISCAP-V; Tissue et al., 2022). Blind to questionnaire responses by parents, clinicians assessed the presence (DISCAP-V diagnosis) and severity of child psychiatric disorder on a scale of 0–6 (DISCAP-V severity rating). Those with a severity rating 4 and above were considered to meet criteria for the specified disorder.

Clinicians also assessed global functioning of the child using the CGAS (Shaffer et al., 1983). CGAS ratings span 0–100; higher scores indicate good health and superior functioning; lower scores indicate illness and impairment. Inter-rater reliability was analysed for a subset of clinical interviews ($n = 24$) using a secondary team of clinicians, who were blind to the primary clinician's ratings. DISCAP-V severity ratings between clinicians correlated highly ($r = .79$, $p <. 001$). CGAS scores moderately correlated ($r = .57$, $p = .004$). Ratings did not significantly differ between clinicians.

Statistical Analysis

To examine the stability of MH symptoms, the first aim, ICCs were assessed for each measure between time points. Listwise deletion was used for missing data.

To examine the concurrent validity of PSC measures with clinician assessments, the second aim, DISCAP-V diagnoses were grouped into four diagnostic categories: internalising (anxious depressive), externalising (disruptive and impulse control) and attention (attention deficit/hyperactivity). Other disorders such as trauma and autism spectrum disorders were grouped into "Other diagnoses". Pearson's correlations were calculated for Preschool PSC total baseline scores, PSC-17 total and subscale baseline scores and severity ratings for each

DISCAP-V diagnostic category, any type of diagnosis and CGAS scores. Pairwise deletion was used for missing data.

To examine predictive validity, baseline PSC scores (Preschool PSC total, PSC-17 total and subscale scores) were classified into at-risk categories using recommended thresholds of ≥ 9 for Preschool PSC total scores (Sheldrick, Henson, Merchant, et al., 2012), ≥ 15 for PSC-17 total scores, ≥ 5 for PSC-17 internalising subscale scores, ≥ 7 for PSC-17 externalising and attention subscale scores (Gardner et al., 1999). Cases were then examined using chi-square tests due to a smaller sample size. Cell counts were sufficient for all analyses.

To examine the difference in level of association between the PSC measures, CBCL and DISCAP-V diagnoses, the third aim, PSC scores were grouped by child age (e.g., 3y.o., 4y.o., 5y.o.) at baseline and correlations calculated. Z-test comparisons for dependent samples were conducted using correlations between the two PSC measures as the third correlation.

5.4 Results

Table 14 presents ICCs for the PSC and CBCL measures between each time point, which were all significant at $p < .001$. ICCs for the Preschool PSC and CBCL total scores both increased across time from .64 to .83, and .67 to .83 respectively. ICCs for PSC-17 total and internalising subscale scores decreased from .67 to .59, and .46 to .39. Externalising and attention subscale scores also decreased but remained relatively stable across time. The stability of PSC-17 scores did not match CBCL subscale ICCs, all of which increased across time. ICC estimates and their 95% confident intervals were calculated based on a single rating, absolute-agreement, 2-way mixed-effects model.

The Preschool PSC and PSC-17 moderately correlated with CGAS functioning scores at the bivariate level, $r = -.59$ and -.49 respectively, $p < .001$, $n = 167$. Correlations between the Preschool PSC and DISCAP-V severity ratings were all significant and ranged from .35–.48, $p < .001$, aside from internalising disorders which were not significantly associated.

Correlations between the PSC-17 total scores and DISCAP-V severity ratings were all significant and ranged from .30–.44, $p < .001$, aside from internalising disorders which were not significantly associated. PSC-17 subscales correlated with DISCAP-V severity ratings for corresponding categories. The PSC-17 externalising subscale correlated with both externalising, $r = .41$, and attention diagnoses, $r = .21$, $p < .001$. Table 15 presents correlations for all scores.

Analysis of at-risk cases as indicated by the PSCs compared to DISCAP-V diagnoses resulted in mixed findings. In terms of predictive validity, the Preschool PSC was highly sensitive to predicting diagnosis, but the PSC-17 was more accurate in terms of specificity across total scores and subscales. Table 16 presents predictive validity results for each PSC measure.

Table 14.

Stability of Screening Scores across Multiple Time Points

ICC (95% CI)	Baseline–3mFU (n = 154)	3mFU–6mFU (n = 140)
PPSC Total score	.64 (.36–.79)**	.83 (.78–.88)**
PSC-17		
Total score	.67 (.56–.75)**	.59 (.13–.79)**
Internalising subscale	.46 (.32–.58)**	.39 (.15–.57)**
Externalising subscale	.55 (.42–.65)**	.54 (.21–.72)**
Attention subscale	.79 (.70–.85)**	.75 (.51–.86)**
CBCL		
Total problems	.67 (.57–.75)**	.83 (.76–.87)**
Internalising subscale	.65 (.55–.73)**	.79 (.72–.84)**
Aggressive syndrome	.69 (.59–.76)**	.80 (.73–.85)**
Attention syndrome	.72 (.63–.79)**	.83 (.77–.88)**

Notes. PPSC = Preschool PSC. **$p < .001$

Table 15.

Pearson's Correlations between PSC Screening Scores, DISCAP-V Severity Ratings and CGAS Scores

	CGAS	Any diagnosis	Internalising diagnosis	Externalising diagnosis	Attention diagnosis	Other diagnosis
PPSC Total score	-.59**	.48**	.15	.45**	.39**	.35**
PSC-17						
Total score	-.49**	.44**	.10	.30**	.41**	.34**
Internalising	-	.32**	.29**	.10	.07	.29**
Externalising	-	.35**	.04	.41**	.21**	.29**
Attention	-	.28**	-.07	.14	.56**	.18*

Notes. $n = 167$. PPSC = Preschool PSC. *$p < .05$, **$p < .001$

Table 16.

Predictive Validity of PSC screening with DISCAP-V Diagnoses

	AUC (CI)	Sensitivity %	Specificity %	PPV	NPV
PPSC Total score > Any diagnosis	.78 (.70–.86)	84.6	54.7	36.26	92.11
PSC-17					
Total score > Any diagnosis	.74 (.65–.82)	15.4	93.0	23.35	78.29
Internalising subscale > Internalising diagnosis	.74 (.63–.84)	16.7	90.9	23.53	86.67
Externalising subscale > Externalising diagnosis	.77 (.60–.94)	50.0	91.7	27.78	96.64
Attention subscale > Attention diagnosis	.82 (.69–.96)	38.5	96.8	50.00	94.90

Notes. $n = 167$. PPSC = Preschool PSC.

Correlations between PSC total scores and DISCAP-V severity ratings were conducted for each age group. For children who were three years of age at baseline, both PSCs moderately correlated with severity ratings (r = .34–.39, p < .001) and strongly correlated with CBCL total problems scores (r = .73–.77, p < .001). For 4-year-old children, both PSCs strongly correlated with both DISCAP-V severity ratings (r = .51–.58, p < .001) and the CBCL (r = .74–.78, p < .001). Correlations for 5-year-old children were insignificant, except for the PSC-17 and CBCL (r = .55, p < .001); however, these results may have been affected by the small number of children in this age range.

For each age, there was no significant difference between the Preschool PSC and PSC-17 and their associations with DISCAP-V severity ratings or the CBCL. Z tests ranged from -0.71–1.26, all p > .05. Results are presented in Table 17.

Table 17.

Association between PSC Screening Scores, DISCAP-V Severity Ratings and CBCL by Age

	Any diagnosis	CBCL
Age 3	(n = 83)	(n = 86)
PPSC	.39**	.77**
PSC-17	.34**	.73**
Z test, significance	z = .60, p = .275	z = .76, p = .223
Age 4	(n = 71)	(n = 78)
PPSC	.58**	.78**
PSC-17	.51**	.74**
Z test, significance	z = 1.26, p = .104	z = 1.02, p = .155
Age 5	(n = 13)	(n = 14)
PPSC	.15	.44
PSC-17	.37	.55*
Z test, significance	z = -1.21, p = .113	z = -0.71, p = .239

Notes. PPSC = Preschool PSC. *p < .05, **p < .001

5.5 Discussion

The purpose of this study was to investigate the stability of SEB symptoms in a preschool community population, examine the validity of the Preschool PSC and PSC-17 compared to clinical assessments, and examine whether concurrent validity for these measures differed by age. To address the first aim of investigating the stability of parent-reported SEB symptoms, ICCs indicated that the Preschool PSC total scores, PSC-17 total scores, externalising and attention subscales all had large and significant measurement stability across time. The PSC-17 internalising subscale had moderate, but still significant, stability across time. The highest levels of stability were found for attention subscales and total scores, which aligns with findings from other studies (Klein et al., 2015).

Comparisons of the stability of screening scores show that this study's findings for the Preschool PSC (ICC .64–.83) were higher than reported ICCs from other studies (.60) (Sheldrick et al., 2025). Findings for PSC-17 scores (ICC .59–.67), however, were lower than other reported ICCs for the same measure (.69) (Sheldrick et al., 2025), which may be a result of differing child age (preschool children in this sample versus older children and adolescents in the previous study) or differences in paediatric populations (community sample in this study versus outpatient paediatric clinics in the previous study). Other studies using alternate brief screening measures, such as the SDQ, reported lower or similar stability across subscales and total problems (.43–.66) compared to our findings for both PSC measures (Klein et al., 2015). It is possible that differences in stability may be due to the longer length of time between assessments reported in other studies (i.e., 21 months) rather than due to the measure per se.

In terms of stability over multiple time points, screening measures displayed mixed findings. Both the Preschool PSC and CBCL showed increasing stability as children aged, perhaps reflecting increasing symptom stability, cognitive development and emotion regulation as children age. Stability can be affected by the psychometric properties of the measure and

measurement error, and by changes in the individual between assessments. The pace of cognitive and physical development occurring at this age may have affected measurement and thus, decreasing stability seen in the PSC-17 may reflect true symptom change rather than reflecting measurement properties. However, if this were the case, we would expect the other measures to also reflect the changing state of the child, which they did not. Further investigation and replication are needed.

The second aim of the study was to investigate the concurrent and predictive validity of PSC measures compared to clinician-rated assessments. Results revealed a strong association of CGAS scores with the Preschool PSC and moderate association with the PSC-17 supporting concurrent validity. The strength of the Preschool PSC correlation with CGAS scores, in which clinicians consider functioning in the home context, may be due to the substantial number of items which target parenting challenges. The current study's findings correspond with previous multi-informant research which found educator-report ratings for the Preschool PSC and the PSC-17 both significantly improved the prediction of clinician-rated functioning scores over and above parent report; however, the incremental validity of educator-report PSC-17 ratings was not as strong as the Preschool PSC (McLean et al., 2025b).

Overall, both Preschool PSC and PSC-17 measures had moderate to strong correlations with the severity of any diagnosis and their corresponding diagnoses, with the exception of internalising disorders. The lack of association between the Preschool PSC and PSC-17 total scores and the severity of internalising diagnoses suggest that the total score may be less related to internalising symptoms in young children. However, PSC-17 subscales, including the internalising subscale, showed significant moderate to strong associations with corresponding diagnoses. Researchers and clinicians screening for internalising disorders in preschool children, therefore, may want to consider using the internalising subscale for the PSC-17 in addition to the total score, or adding a specific measure targeting internalising symptoms.

In terms of predictive validity, the Preschool PSC showed excellent sensitivity, whilst the PSC-17 showed excellent specificity across total and subscale scores. Considered in conjunction with very good AUC values, these results support the ongoing use of measures in early intervention as a screening and triage tool.

Sub-group age analysis was conducted to address the third aim that compared the concurrent validity of the two PSC measures. Results found that the Preschool PSC and PSC-17 did not differ in level of association with the CBCL and diagnosis, at different ages. There was no significant difference between any of the correlations for the Preschool PSC and PSC-17, indicating that there is no specific age at which one measure should be selected over the other. These findings, in fact, support that either measure can be used in the preschool-age range providing users with flexibility to select a measure which targets domains of interest and whether there is a need to measure specific subscales in addition to the total score.

Implications

The first five years of life provide a window for prevention and early intervention before problems become entrenched and more severe, exacerbated by risk factors such as poor parenting practices (Bayer et al., 2012). This study adds to the growing evidence base for MH screening by validating two standardised tools appropriate for use with preschool children that can be used as part of universal screening for early child MH problems. Findings that the PSC measures are strongly associated with clinical diagnoses further justifies their use as a low-cost, first step for universal screening. Current screening and assessment practices in the preschool cohort focus largely on school readiness or literacy and academic competency (Cook et al., 2010; Sim, Thompson, Marryat, Ramparsad, et al., 2019); however, there is an urgent need to increase the scope and screen for SEB difficulties. In doing so, MH screening may be

normalised and destigmatised for parents and caregivers through the process of initiating discussions of early childhood SEB problems and possible pathways for support.

Strengths and Limitations

The strengths of this study include the use of diagnostic interview as the criterion measure in a longitudinal study with a community sample of preschool-age children, one of the few studies to do so. To the best of our knowledge, this is also the first study to investigate the PSC-17 compared to the CGAS, which is a measure widely utilised by clinicians to assess, monitor and guide treatment. One limitation of this study is that it is unknown if families received treatment or other interventions targeting SEB concerns during the study, which may have affected symptom severity at follow-up assessments. Symptom change may be due to accuracy of the measurement or other external factors, such as treatment targeting unwanted behaviour. Following diagnostic interviews, psychologists did not provide detailed feedback to parents and provided only general resources; however, it is possible that families sought further help for their child's SEB wellbeing as a result of the interview. Future research should examine help-seeking behaviour by families during longitudinal research. Future research should also test the robustness of these findings, especially for the stability of the PSC-17 total scores and subscales with a larger sample. Another limitation is the need to investigate the PSC measures' responsiveness or sensitivity to change to increase its clinical utility (Deighton et al., 2014). Sensitivity to change is a psychometric property that is important for outcome measures and clinical practice. Finally, whilst the recruited sample was culturally diverse, parent raters were predominantly female and highly educated. Future research may seek to recruit more diverse parent samples.

Conclusion

Assessing symptom stability in preschool children can be challenging given typical, rapid development; however this study has contributed towards the body of literature investigating the trajectory of child MH in preschool-age children. Shifts in symptoms during the preschool period may indicate the need for more frequent screening during this period so that emerging SEB symptoms are not discounted as a passing phase, but instead, are recognised, identified and directed for further investigation. Regular screening may also mitigate, first, the potential limitations of measures to capture change during a period of rapid developmental change, and second, potential bias from users such as parents or educators who may or may not perceive concerning socio-emotional behaviours. Evidence presented here serves to increase the utility of these measures and enable at-risk children to be identified and to receive early intervention.

CHAPTER 6: Overall Thesis Discussion

6.1 Overview of Findings

The rapid development of children in the preschool years makes it difficult for parents and caregivers to distinguish between social, emotional and behavioural challenges that are normative and transitional, and those which may be indicative of emerging MH problems. This is a key reason why emerging MH problems frequently go unrecognised and therefore untreated in this age range (Huang et al., 2019; Segal et al., 2018). The overall aim of this thesis was to evaluate and validate the PSC screening measures with preschool-age children, using multi-informant reports from parents and educators. As such, this research examined the evidence for existing universal MH screening measures in the preschool-age range and explored the validity of parent- and educator- report versions of the Preschool PSC and PSC-17 for use with young children.

Chapter 2 systematically reviewed the evidence for multi-informant MH screening measures for use by parents and educators of preschool-age children. The review aimed to evaluate the predictive and incremental validity, acceptability and effectiveness of universal screening for child MH problems. Ten measures were found to have been used with parents and educators of preschool-age children; most studies of these measures were conducted within the past decade, indicating that multi-informant, universal MH screening is likely to be an emerging area of research. An evaluation of the psychometric properties for these measures showed strong predictive validity for either parent or educator report, suggesting that the ability of these measures to accurately identify young children at risk of MH problems is promising. However, only one study reported on the incremental validity of parent and educator ratings, finding that when educator ratings are added to parent ratings, there is a significant improvement in the identification of MH outcomes over time. No studies reported on the overall effectiveness of multi-informant screening in terms of rates of identification of MH

difficulties, referral to MH services, or service uptake. Of the three studies examining acceptability, one reported on the acceptability of universal screening programs generally, whilst another two studies reported on the acceptability of a specific screening measure. High acceptability amongst both parents and educators was reported by the one study focussed on screening programs. Two studies focussed on a specific screening measure reported high acceptability amongst parents and moderate acceptability amongst educators.

Overall, the systematic review found few measures could be recommended with full endorsement because few measures meet benchmark standards for developmental screening in terms of predictive validity. The review did, however, identify the PSC-17 as a promising measure, as it was the only freely available measure to be investigated for incremental validity for parents and educators. However, predictive validity for the PSC-17 was not reported in the included study, which highlighted a need to investigate evidence for this measure further. The review found that there is a need for more research into brief, validated multi-informant, MH measures suitable for universal screening with children aged 3–5 years. It highlighted numerous gaps in the existing research, which were the focus of research in subsequent chapters.

Chapters 3–5 addressed a research gap in the literature by investigating the psychometric properties of two promising screening measures for child MH, the Preschool PSC and PSC-17. Chapter 3 aimed to evaluate the reliability, validity and acceptability of the PSC-17 and to produce normative data in an Australian sample of parents. To accomplish this, the study recruited a national sample of Australian parents of preschool-age (3–5 years) and school-age children (6–17 years) to complete measures, including the criterion questionnaire measure, CBCL, at one time point. A subset of participants completed measures at a second time point to examine test-retest reliability. Overall, the PSC-17 was found to have strong internal consistency and excellent concurrent validity with parent-reported CBCL scores across

ages. Of note, the PSC-17 demonstrated strong predictive validity in the preschool-age range, which was surprising, given concerns that it had not been specifically developed or previously validated for this age. The normative data ascertained in this study revealed between 17–25.4% of children and adolescents scored in the "at-risk" range, according to parent reports. The at-risk range means that children scored above recommended clinical thresholds, likely had clinically significant MH problems, and may benefit from receiving early intervention. These proportions were higher than the available normative data from the US and higher than the Australian child and adolescent survey of MH, which identified 13.9% of children (4–17 years) as having a mental disorder (Lawrence et al., 2015). The PSC-17 had high parent acceptability, with 82.9–87.3% of parents indicating that they found the measure to be very or quite appropriate, and there was no significant difference in level of acceptability for this measure between parents of preschool- and school-age children.

Chapter 4 continued the line of research investigating the psychometric properties of the PSC measures through two studies, described in this chapter. The first study aimed to examine the validity, reliability and acceptability of the parent-reported Preschool PSC in young children. Using a national sample of Australian parents of preschool-age children, results supported that the Preschool PSC has excellent internal consistency, adequate test-retest reliability and strong concurrent validity with the CBCL total scores and case classifications. This study also found acceptability of the Preschool PSC amongst parents was very high, ranging from 86.3–96.1% across indices such as appropriateness and clarity. Compared to the findings of Chapter 3, a much higher proportion of children (47.4%) were identified as at risk in this sample when using the existing clinical threshold for the Preschool PSC. Therefore, this study identified an alternative, higher clinical threshold for this measure, which had high sensitivity and specificity, identified fewer at-risk children in the sample, and satisfied benchmark standards for developmental screening measures.

The second study in Chapter 4 aimed to examine the psychometric properties for both the Preschool PSC and PSC-17 in terms of internal consistency, test-retest reliability, inter-rater reliability, predictive, concurrent and incremental validity, and acceptability of educator ratings. Using a paired sample of parents and educators, outcome measures included parent-rated CBCL and clinician-rated diagnosis and functioning. Results supported very high internal consistency, test-retest reliability and small but significant inter-rater reliability between parents and educators for both PSC measures. Results were mixed for the second aim of this study, investigating incremental validity. Incremental validity was not supported for educator ratings when the outcome variable was parent-reported CBCL or clinician-rated diagnosis. This may reflect a bias of the outcome measures towards parent report, or may reflect informant discrepancies between educators, parents and clinicians. Educator ratings significantly improved the prediction of clinician-rated child functioning only when entered into the model before parent ratings on the PSC measures. Using the criterion measure of clinician-rated diagnosis and functioning placed educator-report questionnaires at a disadvantage since clinical interviews were only based on parent reports. Thus, parent-reported questionnaires were more closely aligned with clinician ratings given the overlap in informant, but educator reports may still offer important information about child MH. This study also examined acceptability of the PSC measures amongst educators. Whilst acceptability of both PSC measures was high amongst educators, ranging 78.7–94.7% across indices, educators rated the PSC-17 significantly higher than the Preschool PSC on one acceptability item, concerning the importance of questions asked. Therefore, despite concerns that item wording of the PSC-17 may not be developmentally suitable for the younger age range, results related to the acceptability of these measures did not reflect this. Findings from Chapters 3–4 in relation to acceptability were highly consistent across raters and indicated that both parents and educators

found both PSC screening measures to be highly acceptable, suggesting that either measure was considered appropriate for use with preschool-age children.

After examining the validity of the PSC measures compared to comprehensive questionnaire measures in cross-sectional studies, Chapter 5 extended the earlier investigations by including additional assessment time points in a longitudinal design and adding in-depth clinical assessments conducted by clinicians; in so doing, it addressed some of the limitations of the earlier chapters. Three aims were addressed in this study. First, the study aimed to examine the stability of child MH symptoms using the PSC measures in a community sample of preschool-age children measured over three time points. Second, the study investigated the concurrent and predictive validity of parent-rated Preschool PSC and PSC-17 scores with the best-available criterion measure—diagnostic clinical interview. Third, the study investigated differences in level of association between the Preschool PSC and PSC-17 and outcome measures such as the CBCL and clinical diagnosis, at different ages, in order to ascertain whether the strength of relationship between the two PSC measures and outcome measures differed by child age during the preschool period.

Results from Chapter 5 show that the Preschool PSC total scores, PSC-17 total scores, externalising and attention subscales had significant measurement stability across time, whilst the PSC-17 internalising subscale had moderate, but still significant, stability across time. An interesting finding was the diverging trajectory of stability over time for the two PSC measures. Stability for the Preschool PSC increased over time, whilst the PSC-17 measure showed decreasing stability. Decreasing stability was not shown by the CBCL measure and does not appear to reflect changing symptoms of the child. In fact, the increasing stability shown by the Preschool PSC, CBCL and other studies may reflect improving emotion regulation and cognitive development, which come as a result of natural, maturing child development (Sheldrick et al., 2025). Items in the Preschool PSC, which have been specifically adapted for

preschool-age children, and items targeting the "Parenting Challenges" domain may correspond more closely to the temperament of developing children than the PSC-17, which focusses on disruptive behaviours, anxiety and attention problems (Sheldrick et al., 2025). In addition, the Preschool PSC may show increasing stability due to the Parenting Challenges domain reflecting an increasing ability of parents to manage emotional and behavioural challenges as children age.

Chapter 5 also extended the examination of the PSC measures by validating the parent-reported measures against clinical interviews conducted by clinicians with parents, including ratings of diagnostic disorder, severity of diagnostic disorder and level of child functioning. Both PSC measures had moderate to strong correlations with severity ratings for any diagnostic disorder, although neither Preschool PSC nor PSC-17 total scores were associated with the severity of internalising diagnoses. However, the PSC-17 internalising subscale was significantly associated with diagnosed internalising disorders. These findings using diagnostic outcomes should be considered together with the results from Chapter 3, which found that, with an outcome of predicting CBCL case classifications, the PSC-17 internalising subscale had the lowest sensitivity out of total and subscale scores for the PSC-17 in the preschool-age range, suggesting that severity of internalising symptoms in preschool-age children is not well captured by the total scores of the PSC measures.

The study in Chapter 5 also investigated the association of parent-reported PSC measures with clinician-rated child functioning and found the PSC-17 had a moderate association with functioning scores, whilst the Preschool PSC had a strong association. These results correspond with the earlier finding in Chapter 4, which reported on the incremental validity of parent and educator ratings. The overall variance of parent- and educator-reported ratings was higher for the Preschool PSC than the PSC-17 when the outcome was clinician-rated child functioning. Together, these findings suggest that the Preschool PSC measure

appears to correspond with measures of child functioning better than the PSC-17. This may be due to the Preschool PSC's six items, which target "Parenting Challenges", and may highlight the need for further attention to the child in context with their parents. In terms of parent-report PSC measures and clinical diagnoses, results of predictive validity analysis using clinical diagnoses as the outcome measure showed that the Preschool PSC had excellent sensitivity and the PSC-17 had excellent specificity across total and subscale scores. When considered with the earlier findings of Chapters 3–4 and findings from other studies, the results support the use of both parent-reported PSC measures as a screening tool with preschool-age children, although Preschool PSC had higher sensitivity for predicting any type of diagnosis compared to the PSC-17 total scores.

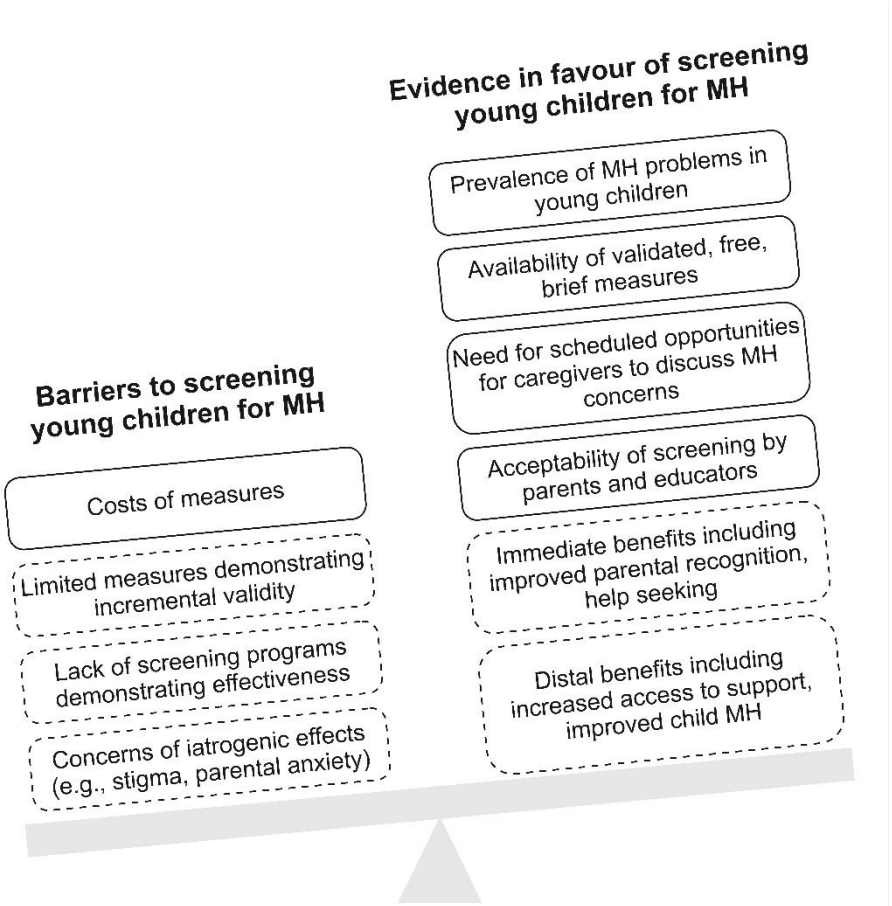
The final aim this study addressed in Chapter 5 was to examine whether the Preschool PSC and PSC-17 differed in level of association with the CBCL and severity of diagnosis at different ages. An examination of the strength of association between PSC measures and severity of clinical diagnoses, and between PSC measures and the CBCL, revealed no significant differences in the strength of association between the two PSC measures at ages 3, 4 or 5 years. This novel finding suggests that there is no specific age at which one PSC measure takes precedence over the other in terms of improved concurrent validity with criterion measures, and that the use of both PSC measures appears to be valid throughout the preschool period from ages 3–5 years.

Overall, this thesis has contributed new research and evidence supporting universal MH screening in young children. There is mounting evidence for MH screening, including the high prevalence of MH problems in young children, which indicates a need to identify children through screening early in the trajectory of problems; availability of suitable measures for screening young children; and the high rates of acceptability of universal screening and screening measures by parents and educators. Further research is required into the effect of

screening programs on parental recognition of child MH problems, rates of help seeking for child MH, and child MH outcomes in the long term. There are a number of areas which, if investigated, could help build the case further for MH screening in young children. However, as outlined throughout the thesis, several barriers to universal screening of young children for MH problems have been raised by researchers and clinicians. These include the costs of measures, which can be prohibitive for large-scale, population-based early intervention programs; the lack of evidence for the validity of screening measures and effectiveness of screening programs; and the possible unintended negative effects of screening, which may affect parents and children. A summary of barriers against and evidence in favour of the implementation of MH screening in young children is presented in Figure 5.

Figure 5.

Evidence For and Against Mental Health Screening in Young Children



Note. Dashed lines indicate areas requiring further research.

6.2 Implications for Research and Clinical Practice

The first five years of life provide a critical window for early intervention before emerging MH problems become entrenched and severe. However, recognising children in need of support is challenging during this period of rapid change. This thesis makes a number of novel contributions to the research literature concerning identification of MH problems in preschool-age children. This includes validating parent-report Preschool PSC and PSC-17 against diagnostic clinical interviews; validating educator-report versions of the PSC measures for use with early childhood educators; ascertaining the acceptability of the PSC measures with both parents and educators of young children; examining the measurement stability of child MH symptoms as captured by the PSC measures in a longitudinal cohort of preschool-age

children in the community; and establishing normative data for the PSC measures in an Australian population—all for the first time, to the best of our knowledge. The work of this thesis thus has important implications for research and clinical practice. Implications focus on the need for routine, universal screening of child MH in the preschool years, the need to train educators to support child MH, and the need for future research to further improve measurement of MH problems in preschool-age children.

6.2.1 Need for Routine, Universal MH Screening for Young Children

Current universal screening and assessment practices in the preschool-age range largely focus on school readiness, literacy and/or academic competency (Cook et al., 2010; Sim, Thompson, Marryat, Ramparsad, et al., 2019); however, there is a need to focus on MH problems, given the high prevalence and negative sequelae of MH problems in the preschool-age range. This research presents new evidence that there are high rates of Australian preschool-age children who score in the at-risk range for MH problems (10–17%) and who may be in need of further assessment and early intervention. This finding highlights the need to increase the scope of current preschool screening practices to include MH difficulties. Findings that the PSC measures completed by parents and educators are strongly associated with longer comprehensive questionnaires and clinical diagnoses adds to the evidence base that brief screening measures might be utilised as a low-cost, first step for universal screening. The work of this thesis provides a clear indication that screening programs are warranted in the preschool-age range, since many young children are at risk for MH problems and therefore in need of early intervention.

Screening should ideally be scheduled at delineated points of transition prior to school entry at age six years, such as at preschool entry and/or exit, or annually on the child's birthday, to align with the age of developmental milestones. If universally mandated, routine screening

for child MH could fill the lengthy gaps between vaccination and developmental milestone checks, for which screening is currently performed in Australia. Lengthy time periods between checks lead to many missed opportunities for identification of and intervention in child MH problems. There is a need to offset parents' perspectives of child MH, which may normalise challenging behaviour during this period, and instead empower parents and educators with evidence-based insights into their child's wellbeing. The stability of MH symptoms during the preschool period, despite this being a period of rapid developmental change, suggests that regular, routine screening may help caregivers to identify at-risk children. Providing caregivers with validated tools at scheduled times may help them to recognise a child's need and access support.

The recommendations to introduce routine, universal MH screening cannot be made without addressing various concerns about language and iatrogenic effects. First, the language of universal screening and associated terms such as developmental monitoring, routine surveillance and MH check need further investigation to ensure they are palatable and acceptable to community members such as parents and educators, and professionals such as medical practitioners and researchers (Carl et al., 2024). Whilst this thesis has utilised the term "screening" as the established term in research literature, recent work focussed on developing child MH screening programs have adopted the terms of "MH check-in" or "wellbeing checks", which have been developed in consultation with researchers, practitioners and community members (Carl et al., 2024; McLellan et al., 2025). Use of alternative terms may help to differentiate screening in the MH field from medical models in which screening refers to the detection of a disease. The development of early identification and screening programs in future should include consultation and co-design with community members and researchers to ensure that terminology around screening, and the process and content of the actual screening program are developed with end users, to ensure high levels of uptake.

Second, universal screening entails possible iatrogenic effects, such as increased parental anxiety, MH stigmatisation and over-medicalisation of MH problems (Frances, 2012). Parental anxiety and stigma may arise when children are positively identified on screening measures and subsequently labelled with having MH problems or disorder, whether appropriately or inappropriately (Dunlevy, 2012; Prior, 2012). Such risks and concerns of unintended effects have previously hampered efforts to implement universal MH screening in Australia with young children in terms of the HKC, although there is no evidence on which to base these concerns. Whilst it has been reported that increased assessment of child MH does not inevitably lead to increased diagnoses and psychotropic prescriptions, and regular assessment can, in fact, connect parents to support and intervention which target concerns (Toumbourou, 2012), research is needed to specifically examine potential iatrogenic effects of universal screening programs. Thus, it is clear that future research needs to investigate levels of parental anxiety, stigma about child MH, and inappropriate use of medication in relation to screening programs.

6.2.2 Training Educators to Conduct Multi-informant Screening in Preschool Settings

The current research found educator ratings have high reliability and support incremental validity for child functioning when entered before parent report. In addition, high rates of acceptability by educators for the PSC measures indicate that educators may be important to target in screening programs for preschool children, and preschools or ECEC services may be appropriate settings for educators to complete screening. Preschools provide an accessible cohort of children and could be integrated into a systematic model of monitoring child MH. Since preschools are less encumbered by issues of stigma and access, routine MH screening in these settings may normalise the process of screening for parents and educators and destigmatise the subsequent parent-educator discussions about MH problems in young

children. However, there is a need for future research to examine the appropriate timing and frequency of screening during the preschool period, to maximise the effectiveness and cost-effectiveness of this approach.

The findings of this thesis warrant further investigation of multi-informant screening of young children that incorporates both parent and educator reports, as it represents the ideal assessment of child MH. Multi-informant screening utilises the insights of caregivers who have deep familiarity with child behaviour and emotions across contexts, captured in validated, evidence-based tools which can then identify children in need of further assessment and support from healthcare professionals. An important finding from this research is that parent reports contributed additional predictive value, over and above that provided by educators, with regards to clinician ratings of functional impairment, which contrasted with the only other known study of incremental validity in this age range. Moore, Dowdy, Fleury, et al. (2022) reported that educator ratings of the PSC-17 added significantly more information over and above parent ratings for an outcome of kindergarten social-emotional readiness. The combination of parent and educator reports through multi-informant screening may increase the validity of identifying children who are at risk of MH problems and need early intervention. However, the paucity of research demonstrating the effectiveness and incremental value of combining parent and educator reports highlights the need for more research in this area. This is particularly pertinent, given the demands of screening on parents' and educators' time. To make the most of caregivers' time, identifying at-risk children must be useful and effective, that is, screening should improve rates of diagnosis, referral or service uptake by parents. Screening may increase the likelihood that parents access support for their child, especially when screening produces recommendations about help seeking and referrals to evidence-based programs and services; however, further research demonstrating this is required. The effectiveness of multi-informant screening needs to be a priority for research in order to

demonstrate that multi-informant reports add value when compared to single-informant screening.

Prior to commencing any universal screening program that includes educators, effective educator training programs are needed. Educators' educational attainment and current levels of training vary widely, from individuals who have completed short vocational courses to others who have acquired university- or college-level qualifications (Stormont & Stebbins, 2005). Regardless of the level of qualifications, educators frequently have little to no specific training in the area of child MH (Stormont & Stebbins, 2005). Studies across countries have identified that educators often feel ill-equipped to identify children with MH disorders such as attention-deficit/hyperactivity disorder or autism spectrum disorder, symptoms of which can first emerge in the preschool-age period (Stormont & Stebbins, 2005; Taresh et al., 2020). Moreover, they lack an understanding of evidence-based research and practice, whilst often being called upon to implement evidence-based interventions, indicating a significant research-to-practice gap (Reinke et al., 2011). Validated multi-informant measures may assist educators to accurately identify children in need of support; however, additional training programs are needed to support educators' knowledge and confidence in screening, and also to help educators communicate about child MH with parents after screening is completed.

Screening may present an important opportunity for parents and educators to discuss child wellbeing; however, such communications also present a risk that educators may inadvertently increase MH stigma for parents, hinder parental help seeking or offer well-meaning, but misplaced, diagnoses. Given low education and workforce entry requirements, in addition to high staff turnover in this sector, training is essential for a workforce that is often unskilled in child MH and inexperienced in communicating with parents about child MH and developmental concerns. Training educators to support appropriate parental help seeking for children is important, because parents are more likely to seek help for their child's MH

problems from a preschool staff member than from medical specialists or GPs (Pavuluri et al., 1996). With appropriate training, educators could refer parents to formal sources of MH support for their child and assist them to progress along the help-seeking pathway. By doing so, educators may increase the likelihood that parents access help, as has been shown previously with school-based interventions (Costello et al., 1998). Moreover, training programs should be easily accessible, low cost, scalable, and delivered in a range of online and in-person formats. The need for training preschool educators in this area is clear, but also presents an opportunity to capitalise on this untapped workforce—a resource that could complement healthcare systems working together to improve child MH and wellbeing.

6.2.3 Measurement Continuity in Clinical and Research Practice

A key implication of the current research is that the PSC measures provide measurement continuity, which involves the use of the same measure across demographic variables, such as child age. The findings of this thesis demonstrate the validity of both the PSC-17 and the Preschool PSC with children aged 3–5 years, and therefore support the use of either measure by clinicians and researchers. This means that either measure can be chosen by researchers and clinicians, depending on the outcomes of interest and relative advantages of each measure. For example, both the parent- and educator-report Preschool PSC demonstrated stronger associations with child functioning compared to the PSC-17. In the longitudinal study, the Preschool PSC also followed the pattern of stability shown by the CBCL, with increasing stability as children aged, whereas the PSC-17 showed decreasing stability over time. This may mean the Preschool PSC is more suited to longitudinal studies; however, both PSC measures demonstrated significant stability between time points.

Alternatively, the PSC-17 may be a preferred measure, due to the ability to assess specific MH domains and utilise the internalising, externalising and attention subscale scores.

In addition, by extending downward the age range with which the PSC-17 can be used, the current research supports the ability of clinicians and researchers to utilise one continuous measure spanning the ages 3–17 years, rather than employing various measures in a piecemeal fashion. Continuity of measures is important in longitudinal research and clinical practice, since it enables the use of the same measure over time to assess stability or changes in child MH. When measures are validated for narrow age ranges, their use is limited, and as such, children can age out of measures during long assessment periods, making longitudinal comparisons difficult. This research, therefore, presents an important research contribution by validating two related measures that enable measurement continuity for child MH.

The current findings may also have implications for clinicians who are seeking to routinely monitor individuals and track their progress over the course of early intervention. A key barrier to regular use of measures for clinicians working in child MH is the availability of measures which are freely available, brief, valid and in digital formats (Tully et al., 2024). Whilst the most commonly used measure of child MH, the SDQ, satisfies some of these criteria, it is not a free digital measure, and the poor predictive validity of the SDQ for the preschool-age range has also been highlighted by the systematic review in Chapter 2 as well as in previous studies (Kersten et al., 2016). The work of this thesis thus serves to provide clinicians with two alternative screening measures, which are both brief, freely available, easily administered and validated for the preschool-age range. Further, the validation of PSC-17 with young children under four years provides clear benefits for research and clinical work conducted with children, as it enables children to be tracked longitudinally and routinely using a single measure.

By validating and extending downward the age range of the freely available PSC measures, this research may lead to increased use of these measures and enable measure harmonisation. There is currently no single measure of child MH and wellbeing that is used across trials by researchers, which means there is a highly fragmented landscape of child MH

research using over 140 measures across Australia (Tsiamis et al., 2024). Given the plethora of measures in use, this is an area ripe for harmonisation. Measure harmonisation is an important research concept in which consistent measures are utilised across studies so that data sets can be combined and aggregated for large-scale comparisons. Consistent outcome measures for child MH research would enable a common language between researchers, improved collaboration and a stronger position from which to advocate for policy change in the child MH sector. Standardising measurement may involve a major upheaval of current practice; however, it is likely to reap transformational benefits across research, clinical and policy sectors.

6.3 Strengths and Limitations

There are a number of strengths to the completed studies as well as several limitations that should be kept in mind when interpreting the findings of this thesis. These strengths and limitations are detailed in this section and include participant characteristics, criterion measures, methodological differences, and access to support during the assessment period.

6.3.1 Participant Characteristics

The recruitment of participants from the community is both a strength and limitation of studies in Chapters 3–5. Firstly, normative samples recruited from the community are advantageous, as they mimic the inclusive, general nature of universal screening, which is designed to be delivered in a non-targeted way at the population level. Minimal eligibility criteria in these studies mitigated against the potential creation of a “supernormal” group—that is, a group that excludes participants with extreme scores, creating an effect of artificially altering the mean score—the risks of which have been outlined elsewhere (Kendall & Sheldrick, 2000). Recruiting from a non-clinical population, however, hampered some of the

intended analyses which would have benefitted from a higher number of clinical cases; in particular, the predictive validity analyses in Chapters 3–4. Whilst much research into the PSC measures has been conducted with clinical samples in the US and elsewhere, future research with Australian samples may focus their efforts on clinical populations to test the robustness of findings.

The use of quotas to recruit the large, national samples for studies in Chapters 3–4 ensured that child gender was evenly represented between male and female children. Parent gender across studies, however, was predominantly female. The samples in Chapters 3–4 (4a) included approximately 30% fathers and male caregivers; however, the longitudinal cohort sample of parents used in Chapters 4–5 was almost exclusively female, with less than 5% identifying as male. This continues a long history of under-representation of fathers in parenting research (Panter-Brick et al., 2014). By testing and validating measures with mothers predominantly, the characteristics of the recruited sample may have omitted findings related to fathers, since previous research has shown that mothers' and fathers' ratings of child behaviour differ for preschool-age children (Davé et al., 2008). Future research should therefore aim to recruit greater proportions of fathers to ensure that findings are generalisable for fathers.

Educators recruited as part of the paired, longitudinal cohort were also predominantly female (97.9%); however, this accurately reflects the composition of the ECEC workforce, which is also predominantly female (97.2–97.6%; Cortis et al., 2023). Future research may explore if there are any characteristics of educators, such as years of experience, strength of relationship with the child, or any other factors amongst educators which predict higher accuracy in identifying child MH problems. Identifying such factors could lay the groundwork for future educator training programs, although we acknowledge the challenges of researching and training a workforce with minimal time and high employee turnover.

6.3.2 Criterion measures

The systematic review in Chapter 2 outlined the important issue of selecting analogous criterion measures, in order to minimise criterion contamination or possible bias when examining the validity of measures. The selection of criterion measures may be a limitation in the subsequent studies. First, common method variance is one limitation that may have affected the findings of Chapters 3–4, as all included assessments, including the criterion measures, were questionnaires. A need for alternate methods of assessment was subsequently addressed in Chapter 5 through the use of clinical interviews conducted by psychologists. However, whilst structured diagnostic interviews are considered the best-available gold standard, these interviews are often based solely on a single informant and this informant is usually a parent rather than an educator. In future, studies investigating the validity of the PSC measures should conduct clinical interviews with large, representative samples and include multiple caregivers to ensure multiple perspectives are included in diagnostic interviews. Alternatively, future research may investigate the predictive power of educator reports in relation to child functioning in a school context. Preschool educator reports may predict later reports from school educators. The incremental validity of parent and educator reports using educator reports as a criterion measure may also reveal new information.

Second, the selection of criterion measures and specifying who rates these measures is important, since this can affect the outcome of comparative analyses which involve different raters. This issue of potential criterion contamination was raised in the review of Chapter 2, in which parent- and educator-rated test measures were compared to educator-rated criterion measures. Educator-rated criterion measures may have biased results towards educator ratings, since they involved the same rater. The studies in Chapters 4 (4b) and 5 addressed this issue of favouring educator reports by including comprehensive parent-rated questionnaire data and clinical interviews with parents. However, neither comprehensive educator-report

questionnaire data nor clinical assessment including interviews with educators were included in these studies. This may have biased results towards parent reports and not sufficiently accounted for MH symptoms displayed outside the home context. Such bias in assessment can be mitigated through multi-informant report that captures perspectives about the child from multiple caregivers and clinicians, and captures diverging perspectives and domain-relevant information occurring across contexts (De Los Reyes et al., 2023). As such, future research should incorporate the ideal assessment of child MH, that is, triangulated criterion measures which include comprehensive parent, educator and clinical assessments in order to validate measures. When investigating the value of multi-informant approaches, it is essential that multi-informant assessment is adopted in all aspects of the study, that is for the reference measures and criterion measures.

6.3.3 Methodological differences

Whilst not a limitation within each standalone chapter, there were differences across studies in methodological approaches taken in relation to predictive validity analyses. Predictive validity analyses can be undertaken alternately through binary logistic regression models or chi-square tests of independence. Both are valid, but offer different approaches with differing purposes, and often the exact methodology selected by researchers is not reported in published articles. This lack of detail in previous studies posed challenges for determining the most appropriate statistical approach. Therefore, various methods were utilised, such that Chapter 3 utilised binary logistic regression models to examine predictive validity, whilst Chapters 4–5 utilised chi-square tests of independence cross-tabulation. Different approaches may lead to differing results in terms of their applicability to a model or specific samples, although this level of detail is frequently not reported in literature. In fact, comparisons of sensitivity and specificity are difficult given these analyses are highly sensitive to the criterion

measure employed in individual studies. Ideally, the methodology would be consistent across studies included in the systematic review of Chapter 2 and within this thesis, and clearly justified within each study. This is a limitation that future research should address.

6.3.4 Access to Support during the Assessment Period

A possible confounding factor that was not controlled for in the outlined studies of Chapters 3–5 was families' access to intervention during the assessment period, which is an important limitation to consider. It is not clear what proportion of families were accessing services at baseline or during the follow-up period. Parents may have accessed treatment during the test-retest period of approximately 35 days for the studies in Chapters 3 and 4(a), or during the three months between assessments as part of the longitudinal study in Chapter 5. Although the longitudinal study of Chapter 5 attempted to control for this by ensuring psychologists did not provide detailed feedback and recommendations to families following clinical interviews at baseline, interviews or participation in the study may have prompted help seeking by parents. Receipt of services by families was not measured and so it is possible that families independently sought out support services between assessment points. Accessing support targeting MH concerns during the studies may have affected symptom severity at follow-up assessments by improving child behaviour or parent management of problems, and thereby reducing perceived burden and improving functioning. Future studies should, therefore, collect information regarding receipt of services to control for whether changes in levels of symptoms over time are due to the accuracy of measures or another external factor, such as intervention for child MH symptoms.

6.4 Directions for Future Research

There are several questions that remain unanswered in the current research and need to be tested in future research studies. First, the findings of this study suggest that screening alone may not adequately identify children with internalising difficulties, and that screening measures for internalising problems in young children need to be developed and refined further. Findings that the total scores for the PSC measures were not associated with the clinician-rated diagnostic severity of internalising disorders and lower sensitivity of the PSC measures when predicting CBCL case classifications for internalising disorders likely reflects the challenges of identifying children with internalising problems in this age range (Duhig et al., 2000; Smith, 2007). For researchers, internalising disorders in preschool-age children have been sidelined due to limitations with available measures, contested conceptualisations of psychiatric diagnosis in preschool children, and a perception that diagnosis in this age range was inappropriate (Egger & Angold, 2006; Morgan et al., 2019). Moreover, reviews of multi-informant assessments have previously found that parents and teachers are more likely to agree about externalising behaviour than they are about internalising concerns (De Los Reyes et al., 2015). Externalising problems are often easier to detect, as they tend to be more clearly observable, demand more attention and disrupt the parent-child relationship, whereas internalising problems can be more easily overlooked in a room full of children vying for educators' attention and have less parental burden (Angold et al., 1998). This is important, given parental burden is a key driver of help seeking by parents. The perception of parental burden and child impairment, in addition to parental recognition of the problem, is what drives subsequent steps to access to MH services (Costello et al., 1998). Future research should, therefore, focus on developing screening measures which can accurately identify internalising problems in this age range.

Second, there are additional psychometric properties that were not tested in the current research on PSC measures. In particular, sensitivity to change and responsiveness were not investigated. Sensitivity to change refers to a measure's capacity to assess symptom change or change in state (Deighton et al., 2014), whilst responsiveness signals the ability of the measure to detect change over time and is considered a form of longitudinal validity (Mokkink et al., 2021). These properties are particularly important for measures involving subjective ratings (e.g., parent or educator report) as opposed to objective, observable measures. In clinical practice, such as routine outcome monitoring, or for research conducted longitudinally, evidence of a measure's sensitivity to change and responsiveness ensures that comparisons of observed scores are not confounded by changes in response shifts, that is, changes in the way an individual responds to assessment (Fokkema et al., 2013). Neither measurement property has been investigated for either the PSC-17 or Preschool PSC. Future research exploring these properties for the PSC measures will likely increase the clinical utility of the measures across research and clinical settings, as they enable statements to be made about the effectiveness of interventions and improvements in treated individual's symptoms. Moreover, researchers may consider the utility of Cronbach's alpha given questions over its ability to estimate reliability and measure internal consistency (Sijtsma, 2009).

Third, future research which examines normative data may consider reporting data in full such that the prevalence of children scoring a particular score on a measure may be deduced from publications. Clinicians may find detailed normative data more useful in clinical settings when discussing the results of screening. Related to the provision of normative data is research on alternative thresholds and unintended negative effects of universal MH screening programs in the preschool-age range, which is needed. Findings in Chapter 4 identified a high proportion of at-risk children using previously established thresholds for the Preschool PSC and highlight the need to consider alternative thresholds in different populations to avoid possible iatrogenic

effects. Alternative thresholds should be considered when screening children in community contexts that differ in prevalence compared to original validation samples or other normative samples, or for contexts that differ in the level of available resources to manage the number of children identified as at risk by screening measures (Pauker & Kassirer, 1980; Sheldrick & Garfinkel, 2017). If universal MH screening is to be implemented at scale, at key points in early childhood, it is imperative that early intervention services for children and their families are available. If services are not available after screening, there may be unexpected costs and risks involved for families and society such as parental stress and worry, wasted resources, and increased healthcare burden for individual practitioners and the wider system (Sheldrick & Garfinkel, 2017). The availability of effective early intervention following screening is an important consideration, since the lack of available services may lead to unintended negative effects of screening programs. However, there is little research in this area, and therefore, this is a priority for future research.

Fourth, there is a scarcity of research investigating the effectiveness of screening programs in early childhood settings. The findings of the review in Chapter 2 showed that there was only one study reporting on the effectiveness of parent-reported screening for an educational outcome. No studies reported on MH outcomes. Whilst school-based MH screening in older children has been demonstrated to be an effective and cost-effective mechanism for managing expensive and inaccessible healthcare services (Humphrey & Wigelsworth, 2016; Murphy et al., 2017), the effectiveness and cost-effectiveness of a universal MH screening program based in ECEC services including preschools has yet to be investigated and is important for making the case for screening of young children (see Figure 5). For the almost 1 million children in Australia aged 3–5 years, routine, universal screening could make a significant impact on improving future child MH trajectories and wellbeing, in addition to ensuring the cost effectiveness of preventative measures delivered at scale

(Australian Bureau of Statistics, 2024). However, there is a need for research to demonstrate the case of effectiveness and cost effectiveness of screening in the preschool-age range, prior to large-scale dissemination.

6.5 Conclusion

To support young children who are unrecognised and untreated for emerging MH problems, there is a need for universal MH screening using brief, validated, multi-informant measures to identify at-risk children in the preschool-age range. The current research contributes new evidence for two screening measures, the Preschool PSC and PSC-17, which can be utilised with parents and educators of preschool-age children. Findings expand the evidence base for these screening measures, presenting evidence that they are valid, acceptable, and appropriate to use with preschool community populations in Australia. Moreover, this research has presented evidence of predictive and incremental validity for these measures, justifying their use with educators. Universal screening measures as part of early intervention programs may help parents to recognise their child's problems, particularly when ratings are informed by other caregivers such as preschool educators. By bridging the gap between recognition of a child's need for support and seeking support from early intervention, screening drives the first stages of the help-seeking process and may help parents to seek support for their child. The PSC measures facilitate the accurate identification of at-risk young children and their use is an important first step to effective early intervention to prevent stable, life-course persistent trajectories of MH problems.

REFERENCES

Achenbach, T. M. (1999). The Child Behavior Checklist and related instruments. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment, 2nd ed.* (Vol. s, pp. 429-466). Lawrence Erlbaum Associates Publishers.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, *85*(6), 1275-1301. https://doi.org/10.1037/0033-2909.85.6.1275

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213-232. https://doi.org/10.1037/0033-2909.101.2.213

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles* (Vol. 30). University of Vermont, Research center for children, youth, & families.

Alakortes, J., Kovaniemi, S., Carter, A. S., Bloigu, R., Moilanen, I. K., & Ebeling, H. E. (2017). Do child healthcare professionals and parents recognize social-emotional and behavioral problems in 1-year-old infants? *European Child & Adolescent Psychiatry*, *26*(4), 481-495. https://doi.org/10.1007/s00787-016-0909-3

Albers, C. A., Glover, T. A., & Kratochwill, T. R. (2007). Where are we, and where do we go now? Universal screening for enhanced educational and mental health outcomes. *Journal of School Psychology*, *45*(2), 257-263. https://doi.org/10.1016/j.jsp.2006.12.003

Alexander, K. E., Brijnath, B., & Mazza, D. (2013). 'Can they really identify mental health problems at the age of three?': Parent and practitioner views about screening young

children's social and emotional development. *Australian & New Zealand Journal of Psychiatry*, *47*(6), 538-545. https://doi.org/10.1177/0004867413476756

Alexander, K. E., Brijnath, B., & Mazza, D. (2015). Parents' decision making and access to preventive healthcare for young children: Applying Andersen's Model. *Health Expectations*, *18*(5), 1256-1269. https://doi.org/https://doi.org/10.1111/hex.12100

Alexander, K. E., & Mazza, D. (2015). Scrapping the Healthy Kids Check: A lost opportunity. *Medical Journal of Australia*, *203*(8), 321-322. https://onlinelibrary.wiley.com/doi/pdfdirect/10.5694/mja15.00683?download=true

Alloush, D., Carl, T., McLean, R. K., & Tully, L. (2025). Parent problem recognition and perceived need for help for child mental health: The role of child age and presenting problem. *[Manuscript submitted for publication]*.

Andersen, R. M. (1995). Revisiting the behavioral model and access to medical care: Does it matter? *Journal of Health and Social Behavior*, *36*(1), 1-10. https://doi.org/10.2307/2137284

Anderson, J. K., Ford, T., Soneson, E., Coon, J. T., Humphrey, A., Rogers, M., Moore, D., Jones, P. B., Clarke, E., & Howarth, E. (2019). A systematic review of effectiveness and cost-effectiveness of school-based identification of children and young people at risk of, or currently experiencing mental health difficulties. *Psychological Medicine*, *49*(1), 9-19. https://doi.org/10.1017/S0033291718002490

Andrews, G. (1999). Efficacy, effectiveness and efficiency in mental health service delivery. *Australian & New Zealand Journal of Psychiatry*, *33*(3), 316-322. https://doi.org/10.1046/j.1440-1614.1999.00581.x

Angold, A., Messer, S. C., Stangl, D., Farmer, E. M., Costello, E. J., & Burns, B. J. (1998). Perceived parental burden and service use for child and adolescent psychiatric

disorders. *American Journal of Public Health*, *88*(1), 75-80. https://doi.org/10.2105/ajph.88.1.75

Australian Bureau of Statistics. (2021a). *Household and families: Census*. https://www.abs.gov.au/statistics/people/people-and-communities/household-andfamilies-census/latest-release

Australian Bureau of Statistics. (2021b). *Preschool education, Australia: 2021*. https://www.abs.gov.au/statistics/people/education/preschool-education-australia/2021

Australian Bureau of Statistics. (2022-23). *Regional population*. https://www.abs.gov.au/statistics/people/population/regional-population/latest-release

Australian Bureau of Statistics. (2023). *New Census insights on income in Australia using administrative data*. Retrieved 2023, May 2 from https://www.abs.gov.au/articles/new-census-insightsincome-australia-using-administrative-data

Australian Bureau of Statistics. (2024). *Australia: Age structure in 2024*. Canberra Retrieved from https://www.abs.gov.au/statistics/people/population/population-clock-pyramid

Australian Government. (2024). *National immunisation program schedule*. https://www.health.gov.au/resources/publications/national-immunisation-program-schedule?language=en

Australian Institute of Health and Welfare (AIHW). (2020). *Australia's children. Cat. no. CWS 69*. Canberra: AIHW

Aylward, G. P. (1997). Conceptual issues in developmental screening and assessment. *Journal of Developmental & Behavioral Pediatrics*, *18*(5), 340-349. https://journals.lww.com/jrnldbp/fulltext/1997/10000/conceptual_issues_in_developmental_screening_and.10.aspx

Barbarin, O., Hitti, A., & Brown, J. (2020). Assessing the severity of concerns about preschool children's self-regulation of attention, behavior, and emotions using the ABLE universal screener: A Rasch analysis. *Journal of Emotional and Behavioral Disorders*, *28*(3), 167-179. https://doi.org/10.1177/1063426619864932

Barbarin, O. A. (2007). Mental health screening of preschool children: Validity and reliability of ABLE. *American Journal of Orthopsychiatry*, *77*(3), 402-418. https://doi.org/10.1037/0002-9432.77.3.402

Bayer, J., Hiscock, H., Scalzo, K., Mathers, M., McDonald, M., Morris, A., Birdseye, J., & Wake, M. (2009). Systematic review of preventive interventions for children's mental health: What would work in Australian contexts? *Australian & New Zealand Journal of Psychiatry*, *43*(8), 695-710. https://doi.org/10.1080/00048670903001893

Bayer, J. K., Ukoumunne, O. C., Mathers, M., Wake, M., Abdi, N., & Hiscock, H. (2012). Development of children's internalising and externalising problems from infancy to five years of age. *Australian & New Zealand Journal of Psychiatry*, *46*(7), 659-668. https://doi.org/10.1177/0004867412450076

Becker-Haimes, E. M., Tabachnick, A. R., Last, B. S., Stewart, R. E., Hasan-Granier, A., & Beidas, R. S. (2020). Evidence base update for brief, free, and accessible youth mental health measures. *Journal of Clinical Child & Adolescent Psychology*, *49*(1), 1-17. https://doi.org/10.1080/15374416.2019.1689824

Brener, S., & Wright, A. (2025). Paediatricians' perspectives surrounding identification and management of early childhood disorders in the United States. *Child: Care, Health and Development*, *51*(1), e70041. https://doi.org/https://doi.org/10.1111/cch.70041

Brinley, S., Tully, L. A., Carl, T., McLean, R. K., Cowan, C. S. M., Hawes, D. J., Dadds, M. R., & Northam, J. C. (2024). Universal child mental health screening for parents: A

systematic review of the evidence. *Prevention Science*, *25*, 798–812. https://doi.org/10.1007/s11121-024-01693-8

Burkett, K., Brown, C. M., Pickler, R., Stanton-Chapman, T., Sharps, P., Jacquez, F., Smith, T., Holland, A., Heeman, A., & Froehlich, T. (2025). Missed opportunities for intervening early in preschoolers with developmental concern: Perspectives from Head Start parents, teachers, and healthcare providers. *Journal of Developmental and Behavioral Pediatrics*, *46*(1), e45-e55. https://doi.org/10.1097/DBP.0000000000001327

Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, *12*(3), 467-488. https://doi.org/10.1017/s0954579400003114

Carl, T., Tully, L. A., McLean, R. K., Dadds, M., Hawes, D. J., Mihalopoulous, C., Chatterton, M. L., Oberklaid, F., Waters, A. M., Shanley, D., Yap, M. B. H., Cann, W. G., Carlick, T., & Northam, J. C. (2024). Increasing parent help-seeking for child mental health: A study protocol for the Growing Minds Check-In, an online universal screening tool. *Contemporary Clinical Trials*, *149*, 107801. https://doi.org/https://doi.org/10.1016/j.cct.2024.107801

Casale, G., & Reyes, A. d. l. (2023). An integrated model for school-based mental health assessment in inclusive education. *Empirische Sonderpädagogik*, *15*(3), 214-233.

Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders: Longitudinal evidence from a birth cohort. *Archives of General Psychiatry*, *53*(11), 1033-1039. https://doi.org/10.1001/archpsyc.1996.01830110071009

Chaffin, M., Campbell, C., Whitworth, D. N., Gillaspy, S. R., Bard, D., Bonner, B. L., & Wolraich, M. L. (2017). Accuracy of a pediatric behavioral health screener to detect

untreated behavioral health problems in primary care settings. *Clinical Pediatrics*, *56*(5), 427-434. https://doi.org/10.1177/0009922816678412

Chau, T., Tiego, J., Brown, L., Coghill, D., Jobson, L., Montgomery, A., Murrup-Stewart, C., Sciberras, E., Silk, T. J., Spencer-Smith, M., Stefanac, N., Sullivan, D. P., & Bellgrove, M. A. (2023). Against the use of the Strengths and Difficulties Questionnaire for Aboriginal and Torres Strait Islander children aged 2–15 years. *Australian & New Zealand Journal of Psychiatry*, *57*(10), 1343-1358. https://doi.org/10.1177/00048674231161504

Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Constructing a roadmap for future universal screening research beyond academics. *Assessment for Effective Intervention*, *35*(4), 197-205. https://doi.org/10.1177/1534508410379842

Cortis, N., Naidoo, Y., Wong, M., & Bradbury, B. (2023). *Gender-based occupational segregation: A national data profile*. Sydney: UNSW Social Policy Research Centre

Costello, E. J., Pescosolido, B. A., Angold, A., & Burns, B. J. (1998). A family network-based model of access to child mental health services. In J. P. Morrissey (Ed.), *Social networks and mental illness* (Vol. 9, pp. 165-190).

Côté, S. M., Boivin, M., Liu, X., Nagin, D. S., Zoccolillo, M., & Tremblay, R. E. (2009). Depression and anxiety symptoms: Onset, developmental course and risk factors during early childhood. *Journal of Child Psychology and Psychiatry*, *50*(10), 1201-1208. https://doi.org/https://doi.org/10.1111/j.1469-7610.2009.02099.x

Croft, S., Stride, C., Maughan, B., & Rowe, R. (2015). Validity of the Strengths and Difficulties Questionnaire in preschool-aged children. *Pediatrics*, *135*(5), e1210-e1219. https://doi.org/10.1542/peds.2014-2920

D'Souza, S., Waldie, K. E., Peterson, E. R., Underwood, L., & Morton, S. M. B. (2017). Psychometric properties and normative data for the preschool Strengths and

Difficulties Questionnaire in two-year-old children. *Journal of Abnormal Child Psychology*, *45*(2), 345-357. https://doi.org/10.1007/s10802-016-0176-2

Davé, S., Nazareth, I., Senior, R., & Sherr, L. (2008). A comparison of father and mother report of child behaviour on the Strengths and Difficulties Questionnaire. *Child Psychiatry and Human Development*, *39*(4), 399-413. https://doi.org/10.1007/s10578-008-0097-6

de Girolamo, G., Dagani, J., Purcell, R., Cocchi, A., & McGorry, P. D. (2012). Age of onset of mental disorders and use of mental health services: Needs, opportunities and obstacles. *Epidemiology and Psychiatric Sciences*, *21*(1), 47-57. https://doi.org/10.1017/S2045796011000746

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858-900. https://doi.org/10.1037/a0038498

De Los Reyes, A., & Langer, D. A. (2018). Assessment and the Journal of Clinical Child and Adolescent Psychology's evidence base updates series: Evaluating the tools for gathering evidence. *Journal of Clinical Child & Adolescent Psychology*, *47*(3), 357-365. https://doi.org/10.1080/15374416.2018.1458314

De Los Reyes, A., Wang, M., Lerner, M. D., Makol, B. A., Fitzpatrick, O. M., & Weisz, J. R. (2023). The Operations Triad Model and youth mental health assessments: Catalyzing a paradigm shift in measurement validation. *Journal of Clinical Child & Adolescent Psychology*, *52*(1), 19-54. https://doi.org/10.1080/15374416.2022.2111684

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

DeLucia, E. A., Dike, J., Andrzejewski, T. M., & McDonnell, C. G. (2022). Autism screening practices in preschools and early childcare centers: A systematic review. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-022-00323-9

Desta, M., Deyessa, N., Fish, I., Maxwell, B., Zerihun, T., Levine, S., Fox, C., Giedd, J., Zelleke, T. G., Alem, A., & F. Garland, A. (2017). Empowering preschool teachers to identify mental health problems: A task-sharing intervention in Ethiopia. *Mind, Brain and Education*, *11*(1), 32-42. https://doi.org/10.1111/mbe.12135

DiStefano, C., Liu, J., & Burgess, Y. (2017). Investigating the structure of the Pediatric Symptoms Checklist in the preschool setting. *Journal of Psychoeducational Assessment*, *35*(5), 494-505. https://doi.org/10.1177/0734282916647648

Doove, B., Feron, J., Feron, F., van Os, J., & Drukker, M. (2019). Validation of short instruments assessing parental and caregivers' perceptions on child health and development for personalized prevention. *Clinical Child Psychology and Psychiatry*, *24*(3), 608-630. https://doi.org/10.1177/1359104518822673

Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice*, *7*(4), 435-453. https://doi.org/10.1093/clipsy.7.4.435

Dunlevy, S. (2012, 12 June). Child health check 'is reckless'. *The Australian*. https://link-gale-com.ezproxy.library.sydney.edu.au/apps/doc/A292667596/AONE?u=usyd&sid=bookmark-AONE&xid=2cc1a226

Eddy, D. M. (1985, January). Screening for cancer in adults. In D. Evered & J. Whelan (Eds.), *Ciba Foundation Symposium 110-The Value of Preventive Medicine* (pp. 88-109). John Wiley & Sons.

Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, *47*(3-4), 313-337. https://doi.org/10.1111/j.1469-7610.2006.01618.x

Eklund, K., & Dowdy, E. (2014). Screening for behavioral and emotional risk versus traditional school identification methods. *School Mental Health*, *6*(1), 40-49. https://doi.org/10.1007/s12310-013-9109-1

Eleni, N. (2021). Preschool educators' views on supporting children with mental health issues in early childhood education. *European Journal of Education Studies*, *8*(3). https://doi.org/10.46827/ejes.v8i3.3633

Ellingson, K. D., Briggs-Gowan, M. J., Carter, A. S., & Horwitz, S. M. (2004). Parent identification of early emerging child behavior problems: Predictors of sharing parental concern with health providers. *Archives of Pediatrics & Adolescent Medicine*, *158*(8), 766-772. https://doi.org/10.1001/archpedi.158.8.766

Ezpeleta, L., Granero, R., la Osa, N. d., Penelo, E., & Domènech, J. M. (2013). Psychometric properties of the Strengths and Difficulties Questionnaire (3–4) in 3-year-old preschoolers. *Comprehensive Psychiatry*, *54*(3), 282-291. https://doi.org/https://doi.org/10.1016/j.comppsych.2012.07.009

Farrelly, T. (2008). The Aboriginal suicide and self-harm help-seeking quandary. *Aboriginal and Islander Health Worker Journal*, *32*(1), 11-15. https://search.informit.org/doi/10.3316/informit.957994991016599

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Feeney-Kettler, K. A., Kettler, R. J., & Dembitzer, L. (2019). Application of a multitrait-multimethod matrix to social, emotional, and behavioral screening. *Journal of Applied School Psychology*, *35*(1), 20-36. https://doi.org/10.1080/15377903.2018.1468843

Feeney-Kettler, K. A., Kratochwill, T. R., Kaiser, A. P., Hemmeter, M. L., & Kettler, R. J. (2010). Screening young children's risk for mental health problems: A review of four measures. *Assessment for Effective Intervention*, *35*(4), 218-230. https://doi.org/10.1177/1534508410380557

Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2011). Identification of preschool children at risk for emotional and behavioral disorders: Development and validation of a universal screening system. *Journal of School Psychology*, *49*(2), 197-216. https://doi.org/https://doi.org/10.1016/j.jsp.2011.02.001

Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment*, *25*(2), 520-531. https://doi.org/10.1037/a0031669

Frances, A. (2012). Better safe than sorry. *Australian & New Zealand Journal of Psychiatry*, *46*(8), 695-696. https://doi.org/10.1177/0004867412453775

Franck, K., Seland, M., Rimul, J., Sivertsen, A. H., & Kernan, M. (2024). Assessing children's psychosocial well-being: Norwegian early childhood education and care teachers' challenges when completing a global screening tool. *Contemporary Issues in Early Childhood*, *25*(4), 445-459. https://doi.org/10.1177/14639491221133454

Gall, G., Pagano, M. E., Desmond, M. S., Perrin, J. M., & Murphy, J. M. (2000). Utility of psychosocial screening at a school-based health center. *Journal of School Health*, *70*(7), 292-298. https://doi.org/10.1111/j.1746-1561.2000.tb07254.x

Gao, R., Raygoza, A., Distefano, C., Greer, F., & Dowdy, E. (2022). Assessing measurement equivalence of PSC-17 across teacher and parent respondents. *School Psychology International*, *43*(5), 477-495. https://doi.org/10.1177/01430343221108874

Gardner, W., Lucas, A., Kolko, D. J., & Campo, J. V. (2007). Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*(5), 611-618. https://doi.org/10.1097/chi.0b013e318032384b

Gardner, W., Murphy, M., Childs, G., Kelleher, K., Pagano, M., & Jellinek, M. (1999). The PSC-17: A brief pediatric symptom checklist with psychosocial problem subscales. *Ambulatory Child Health*, *5*(3), 225-236.

Garton, A. F., Zubrick, S. R., & Silburn, S. R. (1998). The Western Australian Child Health Survey: A review of what was found and what was learned. *The Australian Educational and Developmental Psychologist*, *15*(1), 34-44. https://doi.org/10.1017/S0816512200027838

Girio-Herrera, E., Dvorsky, M. R., & Owens, J. S. (2015). Mental health screening in kindergarten youth: A multistudy examination of the concurrent and diagnostic validity of the Impairment Rating Scale. *Psychological Assessment*, *27*(1), 215-227. https://doi.org/10.1037/a0037787

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*(2), 117-135.

Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of*

Child and Adolescent Psychiatry, *48*(4), 400-403. https://doi.org/10.1097/CHI.0b013e3181985068

Gudmundsson, O. O., Magnusson, P., Saemundsen, E., Lauth, B., Baldursson, G., Skarphedinsson, G., & Fombonne, E. (2013). Psychiatric disorders in an urban sample of preschool children. *Child and Adolescent Mental Health*, *18*(4), 210-217. https://doi.org/10.1111/j.1475-3588.2012.00675.x

Gustafsson, B. M., & Sund Levander, M. (2024). The assessment of preschool children with ESSENCE symptoms: Concordance between parents, preschool teachers and child psychologists. *BMC Pediatrics*, *24*(1), 191. https://doi.org/10.1186/s12887-024-04693-3

Harrison, J. R., Vannest, K. J., & Reynolds, C. R. (2013). Social acceptability of five screening instruments for social, emotional, and behavioral challenges. *Behavioral Disorders*, *38*(3), 171-189.

Harwood, M. D., O'Brien, K. A., Carter, C. G., & Eyberg, S. M. (2008). Mental health services for preschool children in primary care: A survey of maternal attitudes and beliefs. *Journal of Pediatric Psychology*, *34*(7), 760-768. https://doi.org/10.1093/jpepsy/jsn128

Hawes, D. J., Lechowicz, M., Roach, A., Fisher, C., & Dadds, M. R. (2021). Capturing the developmental timing of adverse childhood experiences: The Adverse Life Experiences Scale. *American Psychologist*, *76*(2), 253-267. https://doi.org/10.1037/amp0000760

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, *15*(4), 456.

Heckman, J. J. (2008). The case for investing in disadvantaged young children. *CESifo DICE Report*, *6*(2), 3-8. http://hdl.handle.net/10419/166932

Hiscock, H., Mulraney, M., Efron, D., Freed, G., Coghill, D., Sciberras, E., Warren, H., & Sawyer, M. (2020). Use and predictors of health services among Australian children with mental health problems: A national prospective study. *Australian Journal of Psychology*, *72*(1), 31-40. https://doi.org/10.1111/ajpy.12256

Honda, H., Sasayama, D., Niimi, T., Shimizu, A., Toibana, Y., Kuge, R., Takagi, H., Nakajima, A., Sakatsume, R., Takahashi, M., Heda, T., Nitto, Y., Tsukada, S., & Nishigaki, A. (2023). Awareness of children's developmental problems and sharing of concerns with parents by preschool teachers and childcare workers: The Japanese context. *Child: Care, Health & Development*. https://doi.org/10.1111/cch.13153

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT) version 2018* (Canada Patent No. 1148552). Canadian Intellectual Property Office - Industry Canada. https://content.iospress.com/articles/education-for-information/efi180221

https://journals.sagepub.com/doi/full/10.3233/EFI-180221

Houri, A. K., & Miller, F. G. (2020). A systematic review of universal screeners used to evaluate social-emotional and behavioral aspects of kindergarten readiness. *Early Education and Development*, *31*(5), 653-675.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, L., Hiscock, H., & Dalziel, K. M. (2019). Parents' perception of children's mental health: Seeing the signs but not the problems. *Archives of Disease in Childhood*, *104*(11), 1102. https://doi.org/10.1136/archdischild-2018-315829

Humphrey, N., & Wigelsworth, M. (2016). Making the case for universal school-based mental health screening. *Emotional and Behavioural Difficulties*, *21*(1), 22-42. https://doi.org/10.1080/13632752.2015.1120051

IBM SPSS. (2023). *Statistical Package for the Social Sciences*. In (Version 29.0.1.0)

Jacobson, J. H., Pullmann, M. D., Parker, E. M., & Kerns, S. E. U. (2019). Measurement based care in child welfare-involved children and youth: Reliability and validity of the PSC-17. *Child Psychiatry & Human Development*, *50*(2), 332-345. https://doi.org/10.1007/s10578-018-0845-1

Jellinek, M. S., & Murphy, J. M. (1988). Screening for psychosocial disorders in pediatric practice. *American Journal of Diseases of Children*, *142*(11), 1153-1157. https://doi.org/10.1001/archpedi.1988.02150110031013

Jellinek, M. S., Murphy, J. M., Robinson, J., Feins, A., Lamb, S., & Fenton, T. (1988). Pediatric Symptom Checklist: Screening school-age children for psychosocial dysfunction. *The Journal of Pediatrics*, *112*(2), 201-209.

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, *105*(11), 2283-2290. https://doi.org/10.2105/AJPH.2015.302630

Jureidini, J., & Raven, M. (2012). Healthy Kids Check: Lack of transparency and misplaced faith in the benefits of screening. *Australian & New Zealand Journal of Psychiatry*, *46*(10), 924-927. https://doi.org/10.1177/0004867412460731

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and*

Psychological Measurement, *67*(2), 342-356. https://doi.org/10.1177/0013164407067002 1001

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*, *159*(9), 1548-1555. https://doi.org/10.1176/appi.ajp.159.9.1548

Kendall, P. C., & Sheldrick, R. C. (2000). Normative data for normative comparisons. *Journal of Consulting and Clinical Psychology*, *68*(5), 767-773. https://doi.org/10.1037/0022-006X.68.5.767

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development*, *40*(1), 64-75. https://doi.org/10.1177/0165025415570647

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Üstün, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, *20*(4), 359-364. https://doi.org/10.1097/YCO.0b013e32816ebc8c

Kettler, R. J., Feeney-Kettler, K. A., & Dembitzer, L. (2017). Social, emotional, and behavioral screening: A comparison of two measures and two methods across informants. *Journal of School Psychology*, *64*, 93-108. https://doi.org/https://dx.doi.org/10.1016/j.jsp.2017.05.002

Klein, A. M., Otto, Y., Fuchs, S., Reibiger, I., & von Klitzing, K. (2015). A prospective study of behavioral and emotional symptoms in preschoolers. *European Child & Adolescent Psychiatry*, *24*(3), 291-299. https://doi.org/10.1007/s00787-014-0575-2

Konold, T. R., Walthall, J. C., & Pianta, R. C. (2004). The behavior of child behavior ratings: Measurement structure of the Child Behavior Checklist across time, informants, and

child gender. *Behavioral Disorders*, *29*(4), 372-383. https://doi.org/10.1177/019874290402900405

Kowalenko, N. M. (2012). The three year old social and emotional well-being check is good news. *Australian & New Zealand Journal of Psychiatry*, *46*(10), 928-931. https://doi.org/10.1177/0004867412461534

Kremer, P., de Silva, A., Cleary, J., Santoro, G., Weston, K., Steele, E., Nolan, T., & Waters, E. (2015). Normative data for the Strengths and Difficulties Questionnaire for young children in Australia. *Journal of Paediatrics and Child Health*, *51*(10), 970-975. https://doi.org/https://doi.org/10.1111/jpc.12897

Lavigne, J. V., Lebailly, S. A., Hopkins, J., Gouze, K. R., & Binns, H. J. (2009). The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *Journal of Clinical Child & Adolescent Psychology*, *38*(3), 315-328. https://doi.org/10.1080/15374410902851382

Lavigne, J. V., Meyers, K. M., & Feldman, M. (2016). Systematic review: Classification accuracy of behavioral screening measures for use in integrated primary care settings. *Journal of Pediatric Psychology*, *41*(10), 1091-1109. https://doi.org/10.1093/jpepsy/jsw049

Lawrence, D., Johnson, S., Hafekost, J., Boterhoven de Haan, K., Sawyer, M., Ainley, J., & Zubrick, S. R. (2015). *The mental health of children and adolescents: Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing*. (ISBN: 978-1-76007-188-2). Canberra, ACT: Department of Health

Leiva, L., Antivilo-Bruna, A., Peña, F., Scquicciarini, A. M., Simonsohn, A., Perrin, E., & Murphy, J. M. (2024). Validation of a measurement model to identify socio-emotional difficulties in preschool children: The Preschool Pediatric Symptom Checklist—

Chilean Version. *Child Psychiatry & Human Development*. https://doi.org/10.1007/s10578-024-01714-1

Lim, C., Currie, G. E., Waddington, C. S., Wu, Y., Setijo, S., Leask, J., Marsh, J. A., & Snelling, T. L. (2019). Identification of the determinants of incomplete vaccination in Australian children. *Vaccine: X*, *1*, 100010. https://doi.org/https://doi.org/10.1016/j.jvacx.2019.100010

Lipkin, P. H., & Macias, M. M. (2020). Promoting optimal development: Identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics*, *145*(1). https://doi.org/10.1542/peds.2019-3449

Liu, J., Burgess, Y., DiStefano, C., Pan, F., & Jiang, N. (2020). Validating the Pediatric Symptoms Checklist–17 in the preschool environment. *Journal of Psychoeducational Assessment*, *38*(4), 460-474. https://doi.org/10.1177/0734282919828234

Logan, D. E., & King, C. A. (2001). Parental facilitation of adolescent mental health service utilization: A conceptual and empirical review. *Clinical Psychology: Science and Practice*, *8*(3), 319-333. https://doi.org/10.1093/clipsy.8.3.319

Luo, J., Raat, H., Franse, C. B., Bannink, R., Bai, G., & van Grieken, A. (2022). Correlates of help-seeking by parents for the socioemotional development of their 3-year-old children: A longitudinal study. *BMJ Open*, *12*(1), e052595. https://doi.org/10.1136/bmjopen-2021-052595

Makol, B. A., Youngstrom, E. A., Racz, S. J., Qasmieh, N., Glenn, L. E., & De Los Reyes, A. (2020). Integrating multiple informants' reports: How conceptual and measurement models may address long-standing problems in clinical decision-making. *Clinical Psychological Science*, *8*(6), 953-970. https://doi.org/10.1177/2167702620924439

Massachusetts General Hospital. (2007–2024). *Pediatric Symptom Checklist*. https://www.massgeneral.org/psychiatry/treatments-and-services/pediatric-symptom-checklist

McLean, R. K., Tully, L. A., Brinley, S. K., Carl, T., Turnell, A., Northam, J. C., & Dadds, M. R. (2024). Multi-informant universal mental health screening for preschool-aged children by parents and educators: A PRISMA systematic review. *Clinical Child and Family Psychology Review*, *28*, 1-21. https://doi.org/10.1007/s10567-024-00506-2

McLean, R. K., Tully, L. A., & Dadds, M. (2025). Reliability, predictive validity and normative data for the Pediatric Symptom Checklist-17 (PSC-17) in a national Australian sample. *Australian & New Zealand Journal of Psychiatry*, *59*(8), 702-712. https://doi.org/https://doi.org/10.1177/00048674251342952

McLean, R. K., Tully, L. A., & Dadds, M. R. (2025b). Multi-informant reports of preschool mental health: Validation of parent and educator reports and normative data for the Preschool Pediatric Symptom Checklist and PSC-17. *Child and Adolescent Psychiatry and Mental Health, 19*(1), 138. https://doi.org/10.1186/s13034-025-00985-3

McLellan, L. F., Kuhnert, R.-L., Burns, J. R., & Rapee, R. M. (2025). School-based screening for mental health problems. *Psychiatric Clinics of North America*. https://doi.org/10.1016/j.psc.2025.06.007

McLeod, B. D., Jensen-Doss, A., Lyon, A. R., Douglas, S., & Beidas, R. S. (2022). To utility and beyond! Specifying and advancing the utility of measurement-based care for youth. *Journal of Clinical Child & Adolescent Psychology*, *51*(4), 375-388.

Meirte, J., Hellemans, N., Anthonissen, M., Denteneer, L., Maertens, K., Moortgat, P., & Van Daele, U. (2020). Benefits and disadvantages of electronic patient-reported outcome measures: Systematic review. *JMIR Perioper Med*, *3*(1), e15588. https://doi.org/10.2196/15588

Mellor, D. (2005). Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist*, *40*(3), 215-222. https://doi.org/10.1080/00050060500243475

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. Oxford University Press. http://ebookcentral.proquest.com/lib/usyd/detail.action?docID=648039

Mokkink, L., Terwee, C., & de Vet, H. (2021). Key concepts in clinical epidemiology: Responsiveness, the longitudinal aspect of validity. *Journal of Clinical Epidemiology*, *140*, 159-162. https://doi.org/10.1016/j.jclinepi.2021.06.002

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737-745. https://doi.org/https://doi.org/10.1016/j.jclinepi.2010.02.006

Moore, S. A., Dowdy, E., Fleury, I., DiStefano, C., & Greer, F. W. (2022). Comparing informants for mental health screening at the preschool level. *School Psychology Review*, *51*(5), 589-608. https://doi.org/https://dx.doi.org/10.1080/2372966x.2020.1841546

Moore, S. A., Dowdy, E., Hinton, T., DiStefano, C., & Greer, F. W. (2022). Moving towards implementation of universal mental health screening by examining attitudes toward school-based practices. *Behavioral Disorders*, *47*(3), 166-175. https://doi.org/https://dx.doi.org/10.1177/0198742920982591

Moreira, R. S., Magalhães, L. D. C., Siqueira, C. M., & Alves, C. R. L. (2019). Adaptação transcultural do instrumento de vigilância do desenvolvimento infantil "Survey of Wellbeing of Young Children (SWYC)" no contexto brasileiro. *Revista brasileira de*

crescimento e desenvolvimento humano, *29*(1), 28-38. https://doi.org/10.7322/jhgd.145001

Morgan, A. J., Tamir, E., Rapee, R. M., Lyneham, H. J., McLellan, L. F., & Bayer, J. K. (2019). Online Assessment of Preschool Anxiety: Description and initial validation of a new diagnostic tool. *Child and Adolescent Mental Health*, *24*(3), 259-265. https://doi.org/10.1111/camh.12324

Mulraney, M., Coghill, D., Bishop, C., Mehmed, Y., Sciberras, E., Sawyer, M., Efron, D., & Hiscock, H. (2021). A systematic review of the persistence of childhood mental health problems into adulthood. *Neuroscience & Biobehavioral Reviews*, *129*, 182-205. https://doi.org/https://doi.org/10.1016/j.neubiorev.2021.07.030

Murphy, J. M., Abel, M. R., Hoover, S., Jellinek, M., & Fazel, M. (2017). Scope, scale, and dose of the world's largest school-based mental health programs. *Harvard Review of Psychiatry*, *25*(5), 218-228. https://doi.org/10.1097/hrp.0000000000000149

Murphy, J. M., Bergmann, P., Chiang, C., Sturner, R., Howard, B., Abel, M. R., & Jellinek, M. (2016). The PSC-17: Subscale scores, reliability, and factor structure in a new national sample. *Pediatrics*, *138*(3). https://doi.org/10.1542/peds.2016-0038

Murphy, J. M., Ichinose, C., Hicks, R. C., Kingdon, D., Crist-Whitzel, J., Jordan, P., Feldman, G., & Jellinek, M. S. (1996). Utility of the Pediatric Symptom Checklist as a psychosocial screen to meet the federal Early and Periodic Screening, Diagnosis, and Treatment (EPSDT) standards: A pilot study. *Journal of Pediatrics*, *129*(6), 864-869. https://doi.org/10.1016/s0022-3476(96)70030-6

National Center for Education Statistics. (2016). *Early childhood program participation survey of the 2016 national household education surveys program*. Retrieved from https://nces.ed.gov/nhes/tables/ECPP_HoursPerWeek_Care.asp

National Center for Education Statistics. (2018). *Digest of education statistics, 2018*. Retrieved from https://nces.ed.gov/programs/digest/d18/tables/dt18_202.10.asp

Oh, E., Mathers, M., Hiscock, H., Wake, M., & Bayer, J. (2015). Professional help seeking for young children with mental health problems. *Australian Journal of Psychology*, *67*(3), 187-195. https://doi.org/10.1111/ajpy.12072

Owens, J. S., Storer, J., Holdaway, A. S., Serrano, V. J., Watabe, Y., Himawan, L. K., Krelko, R. E., Vause, K. J., Girio-Herrera, E., & Andrews, N. (2015). Screening for social, emotional, and behavioral problems at kindergarten entry: Utility and incremental validity of parent report. *School Psychology Review*, *44*(1), 21-40. https://doi.org/10.17105/SPR44-1.21-40

Panter-Brick, C., Burgess, A., Eggerman, M., McAllister, F., Pruett, K., & Leckman, J. F. (2014). Practitioner review: Engaging fathers–recommendations for a game change in parenting interventions based on a systematic review of the global evidence. *Journal of Child Psychology and Psychiatry*, *55*(11), 1187-1212.

Paton, K., Gillam, L., Warren, H., Mulraney, M., Coghill, D., Efron, D., Sawyer, M., & Hiscock, H. (2022). How can the education sector support children's mental health? Views of Australian healthcare clinicians. *PloS one*, *17*(1), e0261827. https://doi.org/10.1371/journal.pone.0261827

Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *The New England Journal of Medicine, 302*(20), 1109-1117. https://doi.org/10.1056/nejm198005153022003

Pavuluri, M. N., Luk, S.-L., & McGee, R. O. B. (1996). Help-seeking for behavior problems by parents of preschool children: A community study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(2), 215-222. https://doi.org/https://doi.org/10.1097/00004583-199602000-00015

Prior, M. (2012). Why screening and treating 3-year-olds for mental health problems is not such a good idea. *Australian & New Zealand Journal of Psychiatry*, *46*(8), 700-701. https://doi.org/10.1177/0004867412454340

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(2), 65-76. https://doi.org/10.1007/s10488-010-0319-7

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, *53*(2), 109. https://doi.org/10.1037//0003-066x.53.2.109

Reardon, T., Harvey, K., Baranowska, M., O'Brien, D., Smith, L., & Creswell, C. (2017). What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? A systematic review of qualitative and quantitative studies. *European Child & Adolescent Psychiatry*, *26*(6), 623-647. https://doi.org/10.1007/s00787-016-0930-6

Reardon, T., Spence, S. H., Hesse, J., Shakir, A., & Creswell, C. (2018). Identifying children with anxiety disorders using brief versions of the Spence Children's Anxiety Scale for children, parents, and teachers. *Psychological Assessment, 30*(10), 1342-1355. https://doi.org/10.1037/pas0000570

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, *26*(1), 1-13. https://doi.org/10.1037/a0022714

Rhodes, A. (2017). *Child mental health problems: Can parents spot the signs?* The Royal Children's Hospital. https://www.rchpoll.org.au/polls/child-mental-health-problemscan-parents-spot-the-signs/

Rickwood, D., Thomas, K., & Bradford, S. (2012). *Review of help-seeking measures in mental health: A rapid review* (Evidence Check brokered by the Sax Institute for Beyond Blue). https://www.saxinstitute.org.au/wp-content/uploads/02_Help-seeking-measures-in-mental-health.pdf

Robinson, M., Oddy, W. H., Li, J., Kendall, G. E., De Klerk, N. H., Silburn, S. R., Zubrick, S. R., Newnham, J. P., Stanley, F. J., & Mattes, E. (2008). Pre- and postnatal influences on preschool mental health: A large-scale cohort study. *Journal of Child Psychology and Psychiatry*, *49*(10), 1118-1128. https://doi.org/10.1111/j.1469-7610.2008.01955.x

Rocha, B., & Nunes, C. (2021). Psychometric characteristics of the Portuguese version of the Preschool Pediatric Symptom Checklist for children aged 18 to 60 months. *Psicologia*, *35*(2), 109-118. https://doi.org/10.17575/psicologia.v35i2.1767

Sawyer, A. C. P., Chittleborough, C. R., Lynch, J. W., Baghurst, P., Mittinty, M. N., Kaim, A. L. E., & Sawyer, M. G. (2014). Can screening 4–5 year olds accurately identify children who will have teacher-reported mental health problems when children are aged 6–7 years? *Australian & New Zealand Journal of Psychiatry*, *48*(6), 554-563. https://doi.org/10.1177/0004867413514491

Sawyer, M. G., Arney, F. M., Baghurst, P. A., Clark, J. J., Graetz, B. W., Kosky, R. J., Nurcombe, B., Patton, G. C., Prior, M. R., Raphael, B., Rey, J. M., Whaites, L. C., & Zubrick, S. R. (2001). The mental health of young people in Australia: Key findings from the child and adolescent component of the National Survey of Mental Health and Well-Being. *Australian & New Zealand Journal of Psychiatry*, *35*(6), 806-814. https://doi.org/10.1046/j.1440-1614.2001.00964.x

Sawyer, M. G., Reece, C. E., Sawyer, A. C. P., Johnson, S. E., & Lawrence, D. (2018). Has the prevalence of child and adolescent mental disorders in Australia changed between

1998 and 2013 to 2014? *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(5), 343-350.e345. https://doi.org/https://doi.org/10.1016/j.jaac.2018.02.012

Schnyder, N., Sawyer, M. G., Lawrence, D., Panczak, R., Burgess, P., & Harris, M. G. (2020). Barriers to mental health care for Australian children and adolescents in 1998 and 2013–2014. *Australian & New Zealand Journal of Psychiatry*, *54*(10), 1007-1019. https://doi.org/10.1177/0004867420919158

Schubiner, H., Tzelepis, A., Wright, K., & Podany, E. (1994). The clinical utility of the Safe Times Questionnaire. *Journal of Adolescent Health*, *15*(5), 374-382. https://doi.org/https://doi.org/10.1016/1054-139X(94)90260-7

Scott, S., Briskman, J., & Dadds, M. R. (2011). Measuring parenting in community and public health research using brief child and parent reports. *Journal of Child and Family Studies*, *20*, 343-352. https://doi.org/10.1007/s10826-010-9398-z

Segal, L., Guy, S., & Furber, G. (2018). What is the current level of mental health service delivery and expenditure on infants, children, adolescents, and young people in Australia? *Australian & New Zealand Journal of Psychiatry*, *52*(2), 163-172. https://doi.org/10.1177/0004867417717796

Semansky, R. M., Koyanagi, C., & Vandivort-Warren, R. (2003). Behavioral health screening policies in Medicaid programs nationwide. *Psychiatric Services*, *54*(5), 736-739.

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, *40*(11), 1228-1231. https://doi.org/10.1001/archpsyc.1983.01790100074010

Sheldrick, R., Henson, B., Neger, E., Merchant, S., Murphy, M., & Perrin, E. (2012). The Baby Pediatric Symptom Checklist: Development and initial validation of a new social/emotional screening instrument for very young children. *Academic Pediatrics*, *13*. https://doi.org/10.1016/j.acap.2012.08.003

Sheldrick, R., Henson, B. S., Merchant, S., Neger, E. N., Murphy, J. M., & Perrin, E. C. (2012). The Preschool Pediatric Symptom Checklist (PPSC): Development and initial validation of a new social/emotional screening instrument. *Academic Pediatrics*, *12*(5), 456-467. https://doi.org/10.1016/j.acap.2012.06.008

Sheldrick, R. C., Benneyan, J. C., Kiss, I. G., Briggs-Gowan, M. J., Copeland, W., & Carter, A. S. (2015). Thresholds and accuracy in screening tools for early detection of psychopathology. *Journal of Child Psychology and Psychiatry*, *56*(9), 936-948. https://doi.org/10.1111/jcpp.12442

Sheldrick, R. C., Birudavol, J., Hirshfeld-Becker, D. R., Carter, A. S., Perrin, E. C., Dutta, A., Benheim, T. S., Holcomb Scarpelli, J. M., Boudreau, A. A., Jellinek, M. S., & Murphy, J. M. (2025). Longitudinal stability and predictive value of socio/emotional/behavioral screening in an outpatient pediatric network. *Academic Pediatrics*, *25*(3), 102618. https://doi.org/https://doi.org/10.1016/j.acap.2024.102618

Sheldrick, R. C., & Garfinkel, D. (2017). Is a positive developmental-behavioral screening score sufficient to justify referral? A review of evidence and theory. *Academic Pediatrics*, *17*(5), 464-470. https://doi.org/10.1016/j.acap.2017.01.016

Sheldrick, R. C., Merchant, S., & Perrin, E. C. (2011). Identification of developmental-behavioral problems in primary care: A systematic review. *Pediatrics*, *128*(2), 356-363. https://doi.org/10.1542/peds.2010-3261

Sicouri, G., March, S., Pellicano, E., De Young, A. C., Donovan, C. L., Cobham, V. E., Rowe, A., Brett, S., Russell, J. K., Uhlmann, L., & Hudson, J. L. (2023). Mental

health symptoms in children and adolescents during COVID-19 in Australia. *Australian & New Zealand Journal of Psychiatry*, *57*(2), 213-229. https://doi.org/10.1177/00048674221090174

Sijtsma K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sim, F., Thompson, L., Marryat, L., Law, J., & Wilson, P. (2019). Preschool developmental concerns and adjustment in the early school years: Evidence from a Scottish birth cohort. *Child: Care, Health and Development*, *45*(5), 719-736. https://doi.org/10.1111/cch.12695

Sim, F., Thompson, L., Marryat, L., Ramparsad, N., & Wilson, P. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review. *PloS one*, *14*, e0211409. https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0211409&type=printable

Smith, S. R. (2007). Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *Journal of Psychoeducational Assessment*, *25*(2), 139-149. https://doi.org/10.1177/0734282906296233

Soneson, E., Childs-Fegredo, J., Anderson, J. K., Stochl, J., Fazel, M., Ford, T., Humphrey, A., Jones, P. B., & Howarth, E. (2018). Acceptability of screening for mental health difficulties in primary schools: A survey of UK parents. *BMC public health*, *18*(1), 1404. https://doi.org/10.1186/s12889-018-6279-7

Spence, S. H., & Rapee, R. (1999). *Spence Children's Anxiety Scale*. https://www.scaswebsite.com/1_1_.html

Spencer, A. E., Baul, T. D., Sikov, J., Adams, W. G., Tripodis, Y., Buonocore, O., Jellinek, M., Murphy, J., & Garg, A. (2020). The relationship between social risks and the

mental health of school-age children in primary care. *Academic Pediatrics*, *20*(2), 208-215. https://doi.org/10.1016/j.acap.2019.11.006

Spitzer, R. L., Kroenke, K., Williams, J. B., & the Patient Health Questionnaire Primary Care Study Group. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA, 282*(18), 1737-1744.

Steed, E., & Banerjee, R. (2016). Assessment and early identification of young children with social emotional difficulties and behavioral challenges. *Journal of Intellectual Disability - Diagnosis and Treatment*, *3*, 198-204. https://doi.org/10.6000/2292-2598.2015.03.04.5

Ştefan, C. A., & Miclea, M. (2017). Reliability and validity of two brief screening measures of preschoolers' social–emotional competencies. *School Mental Health*, *9*(1), 44-65. https://doi.org/10.1007/s12310-016-9200-5

Steinbaum, D. P., Chemtob, C., Boscarino, J. A., & Laraque, D. (2008). Use of a psychosocial screen to detect children with symptoms of posttraumatic stress disorder: An exploratory study. *Ambulatory Pediatrics*, *8*(1), 32-35. https://doi.org/https://doi.org/10.1016/j.ambp.2007.08.007

Stoppelbein, L., Greening, L., Moll, G., Jordan, S., & Suozzi, A. (2012). Factor analyses of the Pediatric Symptom Checklist-17 with African-American and Caucasian pediatric populations. *Journal of Pediatric Psychology*, *37*(3), 348-357. https://doi.org/10.1093/jpepsy/jsr103

Stormont, M., & Stebbins, M. S. (2005). Preschool teachers' knowledge, opinions, and educational experiences with Attention Deficit/Hyperactivity Disorder. *Teacher Education and Special Education*, *28*(1), 52-61. https://doi.org/10.1177/088840640502800106

Suman, P., Mushtaq, I., & Tandon, R. (2022). Validation of Hindi translation of Survey of Well-Being of Young Children tool in Indian children. *Indian Pediatrics*, *59*(7), 539-542. https://doi.org/10.1007/s13312-022-2556-8

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (Sixth edition ed.). Pearson Education Limited.

Takayanagi, N., Yoshida, S., Yasuda, S., Adachi, M., Kaneda-Osato, A., Tanaka, M., Masuda, T., Kuribayashi, M., Saito, M., & Nakamura, K. (2016). Psychometric properties of the Japanese ADHD-RS in preschool children. *Research in developmental disabilities*, *55*, 268-278. https://doi.org/https://dx.doi.org/10.1016/j.ridd.2016.05.002

Taresh, S. M., Aniza Ahmad, N., Roslan, S., Ma'rof, A. M., & Mohammed Zaid, S. (2020). Mainstream preschool teachers' skills at identifying and referring children with Autism Spectrum Disorder (ASD). *International Journal of Environmental Research and Public Health*, *17*(12), 4284. https://www.mdpi.com/1660-4601/17/12/4284

Thomas, R., Doust, J. A., Vasan, K., Rajapakse, B., McGregor, L., Ackermann, E., & Del Mar, C. B. (2014). Identified health concerns and changes in management resulting from the Healthy Kids Check in two Queensland practices. *Medical Journal of Australia*, *201*(7), 404-408. https://doi.org/https://doi.org/10.5694/mja14.00637

Tissue, A. D., Hawes, D. J., Lechowicz, M. E., & Dadds, M. R. (2022). Reliability and validity of the DSM-5 Diagnostic Interview Schedule for Children, Adolescents, and Parents-5 in externalizing disorders and common comorbidities. *Clinical Child Psychology and Psychiatry*, *27*(3), 870-881. https://doi.org/10.1177/13591045211061800

Toumbourou, J. W. (2012). Would a universal check of 3-year-olds prevent or create childhood mental health problems? *Australian & New Zealand Journal of Psychiatry*, *46*(8), 703-703. https://doi.org/10.1177/0004867412454587

Tremblay, R. E., Nagin, D. S., Séguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M., Pérusse, D., & Japel, C. (2004). Physical aggression during early childhood: Trajectories and predictors. *Pediatrics*, *114*(1), e43-50. https://doi.org/10.1542/peds.114.1.e43

Tsiamis, E., Jorm, A., Neelakantan, L., Lyons, J., & Morgan, A. (2024). Clinical trials of mental health interventions for Australian children: A rapid review. *[Manuscript submitted for publication]*.

Tully, L. A., Kan, J., Turnell, A., McLean, R., Nowland, T., Liew, O., McFarlane, L., Hawes, D. J., & Dadds, M. R. (2024). Use of measures and measurement-based care in child and youth mental health: A survey of Australian practitioners. *Australian Journal of Psychology*, *76*(1). https://doi.org/https://doi.org/10.1080/00049530.2024.2426662

UK National Screening Committee. (2005). *Child health sub-group report on developmental and behavioural problems*. https://view-health-screening-recommendations.service.gov.uk/document/469/download

Vaz, S., Cordier, R., Boyes, M., Parsons, R., Joosten, A., Ciccarelli, M., Falkmer, M., & Falkmer, T. (2016). Is Using the Strengths and Difficulties Questionnaire in a community sample the optimal way to assess mental health functioning? *PloS one*, *11*(1), e0144039-e0144039. https://doi.org/10.1371/journal.pone.0144039

Verhulst, F. C., Koot, H. M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. *Journal of Abnormal Child Psychology*, *22*(5), 531-546. https://doi.org/10.1007/bf02168936

von der Embse, N., & De Los Reyes, A. (2024). Advancing equity in access to school mental health through multiple informant decision-making. *Journal of School Psychology*, *104*, 101310. https://doi.org/https://doi.org/10.1016/j.jsp.2024.101310

Wagner, J. L., Guilfoyle, S. M., Rausch, J., & Modi, A. C. (2015). Psychometric validation of the Pediatric Symptom Checklist-17 in a pediatric population with epilepsy: A methods study. *Epilepsy & Behavior*, *51*, 112-116. https://doi.org/10.1016/j.yebeh.2015.06.027

Wakschlag, L. S., Briggs-Gowan, M. J., Carter, A. S., Hill, C., Danis, B., Keenan, K., McCarthy, K. J., & Leventhal, B. L. (2007). A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *Journal of Child Psychology and Psychiatry*, *48*(10), 976-987. https://acamh.onlinelibrary.wiley.com/doi/pdfdirect/10.1111/j.1469-7610.2007.01786.x?download=true

Westrupp, E. M., Bennett, C., Berkowitz, T., Youssef, G. J., Toumbourou, J. W., Tucker, R., Andrews, F. J., Evans, S., Teague, S. J., Karantzas, G. C., Melvin, G. M., Olsson, C., Macdonald, J. A., Greenwood, C. J., Mikocka-Walus, A., Hutchinson, D., Fuller-Tyszkiewicz, M., Stokes, M. A., Olive, L., . . . Sciberras, E. (2023). Child, parent, and family mental health and functioning in Australia during COVID-19: Comparison to pre-pandemic data. *European Child & Adolescent Psychiatry*, *32*(2), 317-330. https://doi.org/10.1007/s00787-021-01861-z

Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. *Sociological Methodology*, *8*, 84-136. https://doi.org/10.2307/270754

Wilson, J. M. G., & Jungner, G. (1968). *Principles and practice of screening for disease*. World Health Organization.

Wilson, P., Wood, R., Lykke, K., Hauskov Graungaard, A., Ertmann, R. K., Andersen, M. K., Haavet, O. R., Lagerløv, P., Abildsnes, E., Dahli, M. P., Mäkelä, M., Varinen, A., & Hietanen, M. (2018). International variation in programmes for assessment of children's neurodevelopment in the community: Understanding disparate approaches to evaluation of motor, social, emotional, behavioural and cognitive function. *Scandinavian Journal of Public Health*, *46*(8), 805-816. https://doi.org/10.1177/1403494818772211

Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M. L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, *24*(4), 331-363.

APPENDICES (A) – Supplementary Material

Appendix A-1: Codebook Headings

- Year (published)
- Country
- Study design
- Main aim/goal of study
- Sample size
  - Children n =
  - Educators n =
  - Parents n =
- Gender for children (% of sample)
  - Female
  - Male
  - Non-binary
- Age range of children
- Mean age of children
- Study population
- Study setting, number & type
- Mental health domain screened
- Recruitment criteria: inclusion and/or exclusion
- Recruitment strategy/procedure
- Demographic characteristics for children
- Demographic characteristics for parents
- Criterion measure(s) informant, scale, subscales for comparison measures
- Main screening measures and informants
- Main measure(s) purpose
- Administration format
- Main measure(s) construct of interest
- Main measure age range validated with
- Main measure number of items - educator report
- Main measure number of items - parent report
- Main measure response format
- Reliability:
  - Internal consistency - Parent
  - Internal consistency - Educator
  - Test-retest - Parent
  - Test-retest - Educator
  - Inter-rater reliability, correlation / cross-informant agreement
- Validity:
  - Predictive validity/diagnostic statistics
  - Educator-report:
    - Sensitivity
    - Specificity
    - PPV
    - NPV
  - Parent-report:
    - Sensitivity
    - Specificity
    - PPV
    - NPV
  - Any other predictive validity data?
  - Incremental validity
  - Convergent validity with other measures
- Any other parent & educator similarities, differences, and discrepancies?
- Did screening link to intervention or referral?
- Effectiveness data about the screening
- Acceptability data - Parents
- Acceptability data - Educators

Appendix A-2: Psychometric Properties of Measures

Table 18.

Psychometric Properties of Measures

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Convergent validity (Educator, Parent)	Inter-rater reliability	Internal Consistency (Educator, Parent)	Test-retest reliability (Educator, Parent)
Barbarin et al., (2007)	Study 1: 238, 415 Study 2: NR, NR	ABLE (Stage 1: 10, Stage 2: 40)	Educator: NR Parent: ABLE and ORCB: disobedience .12, $p < .05$ Fearfulness .12, $p < .05$ Aggression .17, $p < .001$ Sadness .11, $p < .05$	Study 1: 76% of the time parents and teachers agreed on which children were or were not designated a case (i.e., when either informant reported a concern and endorsed two or more severity items related to that concern.) Study 2: 77% of the time parents and teachers agreed that a child did not have problems.	NR, NR	NR, NR

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Convergent validity (Educator, Parent)	Inter-rater reliability	Internal Consistency (Educator, Parent)	Test-retest reliability (Educator, Parent)
Doove et al., (2019)	294, 329	VAS (3) PEDS-Dutch Version (10)	Educator (baseline, follow-up): Child competence VAS and C-TRF: -.40, -.32, $p < .001$ Child competence VAS and SDQ: -.47, -.32, $p < .001$ PEDS NR Parent (baseline, follow-up): Parenting VAS and CBCL: -.57, -.51, $p < .001$ Parenting VAS and SDQ: -.47, -.45, $p < .001$ Child behaviour VAS and CBCL: -.56, -.57, $p < .001$ Child behaviour VAS and SDQ: -.46, -.49, $p < .001$ PEDS NR	NR	NR, .70	Educator: Child competence VAS: .90, 95% CI = [0.8, 1.0] PEDS NR Parent: Parenting VAS .80, 95% CI = [0.6, 0.9] Child behaviour VAS .90, 95% CI = [0.8, 1.0] PEDS .80, 95% CI = [0.5, 0.9]
Ezpeleta et al., (2013)	Phase 1 N/A, 1,341 Phase 2 94, 622	SDQ (25)	Educator: SDQ and CBCL Int .19 Ext .26 Tot .17 SDQ and DICA-PPYC Int .17 Ext .35 Tot .26 Parent: SDQ and CBCL Int .52	Partial measurement invariance: [$\Delta\chi2(16) = 23.7$; $p = 0.096$], 80% (21 of 25) of the factor loading parameters were equivalent across parents' and teachers' reports	Educator Model 1: .67 (prosocial) - .93 (hyper-activity) Parents Model 2: Int .81 Ext -.91	NR, NR

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Convergent validity (Educator, Parent)	Inter-rater reliability	Internal Consistency (Educator, Parent)	Test-retest reliability (Educator, Parent)
			Ext .62 Tot .58 SDQ and DICA-PPYC Int .41 Ext .56 Tot .54			
Feeney-Kettler et al., (2011)	112, 113	PBSS (46)	NR	PBSS Phase 2 Int *r* = .14 Ext *r* = .52 Prosocial Behavior Scale scores *r* = .31 Tot *r* = .35	Composite: Int .85 (2–3 years); .87 (4–5 years) Ext .87 (2–3 years); .90 (4–5 years)	Educator 9–70 days Int .85 Ext .86 Parent 8–65 days Int .86 Ext .81
Feeney-Kettler et al., (2019)	122, 122	PBSS (46)	NR	PBSS Int *r* = -.09 –.10 Ext *r* = .27–.37 Tot *r* =.26	Educator Int .90 [.87,.92] Ext .97 [.96,.98] Prosocial .96 [.95,.97] Tot .97 [.96,.98] Parent Int .85 [.81,.88] Ext .94 [.92,.95] Prosocial: .90 [.87,.92]	NR, NR

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Convergent validity (Educator, Parent)	Inter-rater reliability	Internal Consistency (Educator, Parent)	Test-retest reliability (Educator, Parent)
					Tot .94 [.92, .95]	
Girio-Herrera et al., (2015)	Study 1: 56, 12 Study 2: 568, 273	IRS (31)	NR	Study 1: Teacher IRS overall and parent-rated Int .03; Teacher IRS overall and parent-rated Ext .25; Teacher IRS overall and parent-rated adaptive skills -.29 Parent IRS overall and teacher-rated Int .07; Parent IRS overall and teacher-rated Ext .16; Parent IRS overall and teacher-rated adaptive skills -.11 Study 2: Parent-rated IRS scores and teacher-rated BESS scores < .24.	NR, NR	NR, NR
Kettler et al., (2017)	105, 105	PBSS (46)	Educator PBSS and BESS Tot .89 PBSS and ASEBA Tot .83, $p < .05$	PBSS Total Score $r = .26$ Nomination rubrics r .37–.40 Rating scales $r = .05$–.36	Educator Int .88 Ext: .96 Prosocial .96 Tot .97	NR, NR

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Convergent validity (Educator, Parent)	Inter-rater reliability	Internal Consistency (Educator, Parent)	Test-retest reliability (Educator, Parent)
			Parent PBSS and BESS Tot .70, $p < .05$ PBSS and ASEBA Tot .61, $p < .05$		Parent Int .87 Ext .94 Prosocial .90 Tot .94	
Moore, Dowdy, Hinton, et al., (2022)	40, 330	BASC-3 BESS (86)	NR	NR	NR, NR	NR, NR
Moore, Dowdy, Fleury, et al., (2022)	14, 535	BASC-3 BESS (86); PSC-17 (17)	Educator PSC-17 and BESS-BERI .89, $p < .001$ Parent PSC-17 and BESS-BERI .68, $p < .001$	BESS $r = .64$ PSC-17 $r = .26$ Tot κ (BESS) = .08 (95% CI [-.02, .19]), $p = .046$ κ (PSC-17) = .09 (95% CI [-.01, .21]), $p = .022$	Educator BASC-3 α = .83 PSC-17 α = .92 Parent BASC-3 α = .87 PSC-17 α = .78	NR, NR
Ştefan et al., (2017)	NR, 180	ECS (30)	Educator ECS/SCS-T and ASEBA scales: -.69– .09 Parent ECS/SCS-P and ASEBA scales: -.67– -.08	ECS-P/ECS-T total and subscale scores: .19–.40 SCS-P/SCS-T total and subscale scores: .20–.27,	Educator ECS-T Emotional competence scale α = .84–.90 SCS-T: α = .92–.94 Parent: ECS-P Emotional competence scale α = .74–.80 SCS-P: social competence scale from the SCS-P, α = .81–.84,	Educator ECS-T: .82–.88 SCS-T: .80–.87 Parent ECS-P: .80–.85 SCS-P: .81–.86.

Author (Year)	Sample *N* (Educator, Parent)	Screening measure (no. of items)	Convergent validity (Educator, Parent)	Inter-rater reliability	Internal Consistency (Educator, Parent)	Test-retest reliability (Educator, Parent)
Takayanagi et al., (2016)	NR, 838	ADHD-Rating Scale-IV Parent (18) and Educator versions (18)	Educator NR Parent ADHD-RS total score and SDQ total difficulties $r = .70$, $p < .001$	ADHD-RS Tot $r = .27$ ($p < .001$)	Educator inattention α = .93 Hyperactivity-impulsivity α = .93 Parent inattention α = .86 Hyperactivity-impulsivity α = .85	NR, NR

Notes. ABLE = Attention Behaviour Language Emotions; ADHD = Attention-Deficit/Hyperactivity Disorder; ASEBA = Achenbach System for Empirically-Based Assessment; BASC-2/3 = Behaviour Assessment System for Children Second/Third edition; BERI = Behavioural and Emotional Risk Index; BESS = Behavioral and Emotional Screening System; CBCL = Child Behavior Checklist; C-TRF = Caregiver-Teacher Rating Form; DICA-PPYC = Diagnostic Interview for Children and Adolescents for Parents of Preschool And Young Children; ECS-P/T= Emotion Competence Screening Parent/Teacher Form; Ext = Externalising; K = Kindergarten; Int = Internalising; IRS = Impairment Rating Scale; NR = Not reported; ORCB = Observer Ratings of Child Behaviour; PBSS = Preschool Behavior Screening System; PEDS = Parents' Evaluation of Developmental Status; PSC-17 = Pediatric Symptom Checklist-17; SCS-P/T = Social Competence Screening Parent/Teacher Form; SDQ = Strengths and Difficulties Questionnaire; TCRS = Teacher Child Rating Scale; Tot = Total scores; VAS = Visual Analogue Scales.

Appendix A-3: Additional Detail about Recruitment Procedure for Study 3

Participants were recruited from an Australian research panel. Panel members were told a 15-minute survey was available for completion. Interested individuals completed screening questions to assess eligibility. Individuals who were eligible were provided with participant information statements, completed consent and completed questionnaires as outlined in the Method. At the conclusion of the questionnaires, participants received a debriefing statement and were then directed to the research panel's website to receive compensation.

Participants voluntarily participated in the anonymous study and received payment upon completion of the study. Payment consisted of online points which could be converted into a gift card once a specified threshold was reached. Digital verification was conducted by the research panel to ensure duplicate responses were not received.

The online questionnaires for parents of preschool-age children (3–5 years) and school-age children (6–17 years) were available for completion from December 2023 to February 2024.

Among the 1700 parents of preschool and 1387 school children who started the questionnaires, 1454 parents of preschool and 1140 school children completed baseline questionnaires. Participants who completed baseline questionnaires were invited to participate in the second assessment. The target sample for the second assessment was 10% of the overall sample, i.e., n = 100 for each age group. The reported sub-sample was the number of participants who completed both the first and second assessments.

To maximise data integrity, data quality checks screened out any respondents who did not report data for a child in the relevant age range and/or failed attention checks. As part of eligibility screening procedures, participants were asked if they had a child in the relevant age range, either preschool- (3–5 years) or school-age (6–17 years). After indicating they had a child in either age range, they proceeded to complete questionnaires including a subsequent

item which asked for the child's date of birth. If the calculated age for the child did not fall into the relevant age range, participants' data were removed. Questionnaires also included attention checks to invalidate responses from speeders or random answering. These items asked respondents to "Select very true to show you are paying attention"; "Very true" being one of the listed response options.

Appendix A-4: Demographic Detail about Participants in Study 3

Preschool parents were aged 36.30 years on average (*SD* = 8.10) and school parents were aged 41.84 years on average (*SD* = 8.34); 66.6% of preschool parents and 68.8% of school parents identified as female; and 84.1% of preschool parents and 78.0% of school parents were married or in *de facto* relationships.

Parent demographic details including marital status, residential location (metropolitan, regional/rural), and household income were compared to available data from the Australian Bureau of Statistics (2021a, 2022–23, 2023). Across the total sample, the majority of parents were married or in a *de facto* relationship (81.1%), which is broadly comparable to the Australian population (91.1% married or in a de facto relationship). In addition, three quarters of parents lived in metropolitan areas (74.4%), compared to 72% of the Australian population (Australian Bureau of Statistics, 2022–23). Six in ten of parents reported annual household incomes of over $100,000 (62.1%), which is also broadly comparable the proportion in the Australian population (56.7% earning over $100,000).

Appendix A-5: Acceptability Measure

Items index confidence in ratings (“Based on your knowledge of the child’s life, how confident are you in the responses/ratings you have given today as part of the Checklist?”); the appropriateness and clarity of item wording (“How appropriate did you find the questions?”; “How clear did you find the wording of the questions?”, and “The Checklist asks important questions about [child]'s mental health.”).

Appendix A-6: Factor Analysis and Model Fit

Model fit was evaluated using the indices of χ^2, comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR). Acceptable model fit has been defined by the following criteria: $\chi 2$ ($p < .05$), RMSEA (≤ 0.06, 90% CI LB ≤ 0.06), SRMR (≤ 0.08), CFI (≥ 0.95), and TLI (≥ 0.95) (Hu & Bentler, 1999). An additional goodness of fit index is the ratio of χ^2/df, however, there is no consensus about an acceptable ratio ranging from ≤ 2.0 to 5.0 (Tabachnick & Fidell, 2014; Wheaton et al., 1977).

To improve the fit of the preschool model, modification indices were examined to determine if additional parameters would improve the model. Correlating three pairs of error terms (items 11 and 17; 13 and 17; and 15 and 16) indicated that the overall chi square would reduce by 122.93. The modified model resulted in a limited improvement in model fit: χ^2 (113, N = 1,048) = 530.995, $p < .001$; $\chi^2/df = 4.70$. RMSEA = .059, CI 90% .054–.065, SRMR = .0491, CFI = .934, TLI = .920.

Modification indices did not indicate that any additional parameters would substantially improve the school-age model.

Factor loadings are presented in Table 19.

Table 19.

Factor Loadings for the PSC-17

Item	Children aged 3–5 years			Children aged 6–17 years		
	Internalising Factor	Externalising Factor	Attention Problems Factor	Internalising Factor	Externalising Factor	Attention Problems Factor
Feels sad, unhappy	.49			.59		
Worries a lot	.59			.65		
Seems to be having less fun	.60			.57		
Feels hopeless	.65			.72		
Is down on him or herself	.70			.74		
Teases others		.55			.59	
Blames others for his or her troubles		.59			.67	
Refuses to share		.63			.60	
Does not understand other people's feelings		.66			.67	
Does not listen to rules		.67			.71	
Fights with other children		.67			.65	
Takes things that do not belong to him or her		.68			.59	
Daydreams too much			.56			.62
Acts as if driven by a motor			.60			.57
Fidgety, unable to sit still			.75			.70
Distracted easily			.82			.84
Has trouble concentrating			.84			.85

Appendix A-7: Additional Detail about Recruitment Procedure for Study 4a

Participants were recruited from an Australian research panel using quotas based on census data such as household income, residential location and parental marital status. Additional quotas applied to child participants such as child gender and age to ensure a balanced sample. Caregivers could include anyone in a caregiving role such as fathers, mothers, kinship carers, or foster carers. Panel members were told a 15-minute survey was available for completion. Interested individuals completed screening questions to assess eligibility. Individuals who were eligible were provided with participant information statements, completed consent and completed questionnaires as outlined in the Method. The online questionnaires for parents of preschool-age children (3–5 years) were available for completion from December 2023 to February 2024.

At the conclusion of the questionnaires, participants received a debriefing statement and were then directed to the research panel's website to receive compensation. Participants voluntarily participated in the anonymous study and received payment upon completion of the study.

Among the 1700 parents of preschool children who started the questionnaires, 1454 parents of preschool children completed baseline questionnaires. Digital verification was conducted by the research panel to ensure duplicate responses were not received. To maximise data integrity, data quality checks screened out any respondents who did not report data for a child in the relevant age range and/or failed attention checks.

The reported sub-sample was the number of participants who completed both the first and second assessments.

Appendix A-8: Demographic Detail about Participants in Study 4a

Table 20.

Child Ethnicity for Sample in Study 4a

Child Ethnicity	*N*	Percent
Caucasian (e.g., British, European)	683	65.3
Other or multiple ethnicities identified	117	11.2
Aboriginal/Torres Strait Islander	76	7.3
South Asian (e.g., Indian, Pakistani)	53	5.1
South-East Asian (e.g., Vietnamese, Filipino)	44	4.2
East Asian (e.g., Chinese, Japanese, Korean)	42	4
Polynesian (e.g., Pacific Islander, Māori)	12	1.1
Middle-Eastern (e.g., Egyptian, Iraqi, Lebanese)	10	1
African/African American	5	0.5
Hispanic/Latino	3	0.3
Total	1,045	100

Appendix A-9: Demographic Detail about Participants in Study 4b

Parents were aged 28–47 years (M = 37.51; SD = 4.29); 95.7% identified as female and 91.5% were married or in *de facto* relationships.

The ethnicity of educators was predominantly Caucasian (75.5%). Remaining ethnicities identified were East Asian (6.4%), South-East Asian (5.3%), South Asian (4.3%), Hispanic/Latino (3.2%), Middle Eastern (1.1%), or multiple ethnicities. No educators identified as Aboriginal or Torres Strait Islander or African/African-American.

APPENDICES (B) – Published Articles

Published peer-reviewed journal articles based on work undertaken for the degree of Doctor of Philosophy.

Clinical Child and Family Psychology Review
https://doi.org/10.1007/s10567-024-00506-2

Multi-Informant Universal Mental Health Screening for Preschool-Aged Children by Parents and Educators: A PRISMA Systematic Review

R. K. McLean[1] · L. A. Tully[1] · S. K. Brinley[2] · T. Carl[1] · A. Turnell[1] · J. C. Northam[1] · M. R. Dadds[1]

Accepted: 20 October 2024

Abstract

Children develop rapidly in the preschool period, making accurate appraisals of mental health (MH) difficult. The preschool years are a key period for early identification of MH concerns and could benefit from multi-informant, universal MH screening (UMHS). This systematic review aimed to identify multi-informant UMHS measures for preschool-aged children, and to examine their clinical utility, effectiveness, and acceptability. Studies reporting the predictive and incremental validity, effectiveness or acceptability of parent and educator-report UMHS measures for children aged 3–5 years were identified through CINAHL, Embase, ERIC, Medline, PsycINFO, Scopus and Web of Science. Studies were excluded if screening was not the primary focus, not universal, single informant, or primarily focussed on Autism Spectrum Disorder. A total of 11 studies using 10 measures was identified. Ten studies screened for broad MH difficulties. Three educator-report and one parent- and educator-report measures had acceptable predictive validity. One study reporting incremental validity found that adding educator-report to parent ratings significantly improved the identification of MH concerns. No studies reported on effectiveness. Three studies that explored acceptability reported strong support for either UMHS in general or specific measures. There are promising results that UMHS can accurately identify child MH concerns in the preschool period using parent and educator reports. However, with few of the examined measures reaching the benchmark standards for predictive validity and only one study examining incremental validity, further research is needed to establish clinical utility. UMHS with preschool populations appears to be acceptable; future studies should further examine multi-informant screening in preschool populations. This systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO; https://www.crd.york.ac.uk/prospero; registration number: CRD 42022383426).

Keywords Child mental health · Preschool · Early learning · Universal screening · Multi-informant · Educator · Teacher · Parent · Universal screening · Universal mental health screening · Socio-emotional and behavioural wellbeing

Children develop rapidly in the first 5 years of life and it is well established that many mental health (MH) difficulties first emerge during the preschool period from 3 to 5 years of age (Bayer et al., 2009; Lavigne et al., 2009; Oh et al., 2015; Robinson et al., 2008). Rapidly changing behaviour, social, and emotional development during this period of childhood can make accurate appraisals of child wellbeing difficult for parents (Konold et al., 2004). Parents and other caregivers of preschool-aged children, such as educators or teachers (hereafter referred to as educators), face challenges in distinguishing between normative behaviour and behaviour indicative of more serious concerns (Wakschlag et al., 2007). For example, behaviours such as tantrums, non-compliance and aggression can be age-appropriate and transitory, but may also be indicative of serious, disruptive behaviour problems, especially when these behaviours are severe and/or persist over time. Thus, primary caregivers such as parents and educators need tools to help them identify children who are at risk of developing MH difficulties, to increase access to early interventions. Universal mental health screening (UMHS) is

✉ M. R. Dadds
mark.dadds@sydney.edu.au

R. K. McLean
rebecca.mclean@sydney.edu.au

[1] School of Psychology, Faculty of Science, University of Sydney, Sydney, NSW, Australia

[2] School of Medicine and Psychology, The Australian National University, Canberra, Australia

Published online: 14 November 2024

a population-based approach in which individuals complete systematic assessments designed to identify those at risk of developing socio-emotional (internalising) and behavioural (externalising) difficulties (Humphrey & Wigelsworth, 2016). This review aims to examine the current literature that has examined UMHS measures in preschool children through multi-informant reports by parents and educators.

Early Intervention and Identification in Children

Early intervention refers to a broad array of academic, medical or social support or treatment programs designed to enhance a young child's development (Ramey & Ramey, 1998). Programs can be delivered to at-risk children, with the aim of being preventative, or targeted treatment programs implemented with children with known or diagnosed problems. Identifying children who may be at risk of MH difficulties, or recognising specific behaviour or emotional difficulties as requiring further investigation is the first step in early intervention. Early identification of children with MH concerns often leads to intervention, thus disrupting the often stable trajectory of MH difficulties from childhood to adulthood.

Parents and other primary caregivers play a significant role in help seeking for their children, that is, the recognition of emerging MH problems and facilitating access to services (Schnyder et al., 2020). The parental help-seeking process has been identified as (1) Parents' recognition of MH difficulties, (2) Parents' recognition of a need for professional help, (3) Parents' actively seeking help, and (4) Family accessing and receiving necessary help/support (Reardon et al., 2017). However, parents do not always recognise behavioural or emotional difficulties as requiring further investigation, nor do they always appreciate the need for early intervention despite the established evidence base (Kowalenko, 2012). This is particularly pertinent in the preschool age range in which differentiating between psychopathology and normal early childhood development is a major challenge (Sim et al., 2019). A nationally representative survey of Australian parents of children under 18 years found that only 35% of parents were confident that they would recognise the signs of MH difficulties in their children and fewer than half (44%) knew where to access professional help if their child was experiencing difficulties, with confidence rates significantly lower in parents of younger than older children (Rhodes, 2017). A longitudinal study has shown that parents who recognise their child's problem behaviour are more likely to access professional help for their child at age 3 (Oh et al., 2015). Thus, this evidence demonstrates the importance of recognising MH concerns as the first step towards parental help-seeking, including further assessment and accessing early intervention services.

Children with untreated MH difficulties can experience far-reaching negative effects throughout adolescence and into adulthood. In children as young as 3, behavioural and emotional difficulties have influenced wellbeing in later childhood and early adulthood across domains such as employment, education, criminal activity, physical and MH (Caspi et al., 1996; Jones et al., 2015). However, left untreated, the implications of low rates of identification mean that children in need of MH services frequently do not receive appropriate intervention and face poor outcomes.

Numerous papers have outlined the poor rates of help-seeking particularly amongst parents of young children (Ellingson et al., 2004), low rates of professional service utilisation amongst preschoolers with clinical diagnoses (Lavigne et al., 2009) or unmet MH needs in children more generally (Hiscock et al., 2020; Kataoka et al., 2002). In fact, the highest rates of unmet MH needs in the United States are in children under 6 years, who are also children of colour and from low-income families (Kataoka et al., 2002).

Universal Mental Health Screening

Given overburdened healthcare systems and the long-term ramifications of untreated MH difficulties, UMHS offers a cost-effective mechanism for identifying young children in need of support (Humphrey & Wigelsworth, 2016). UMHS can be an effective tool for identifying MH concerns by improving recognition and increasing access to early intervention services, thereby facilitating help-seeking by parents or other informants (Humphrey & Wigelsworth, 2016). UMHS provides early identification of individuals with MH symptoms, who might require further follow up and it may also include referral pathways to appropriate services (Lavigne et al., 2016). It can be conducted as a single assessment with questionnaires which are scored according to clinical cut-offs, or involve multistage assessments in which children are identified in a first stage and receive further follow-up or assessment in subsequent stages. As with other standardised assessments, it can be completed individually by children, parents, educators, or other health professionals. UMHS does not make diagnoses of specific disorders and is not a diagnostic tool.

UMHS in the preschool years is important because it provides a critical opportunity to identify children at risk of MH difficulties and provide early intervention prior to difficulties becoming more entrenched and severe (Moore et al., 2022a). This is especially important since early intervention, implemented in early childhood, has demonstrated high rates of cost effectiveness (Heckman, 2008). Despite the potential benefits of UMHS for preschoolers, there is

a paucity of research examining the efficacy, acceptability, and accuracy of UMHS in this developmental period (Anderson et al., 2019). A recent review by Becker-Haimes et al. (2020) of free, brief, and accessible child MH measures identified only two UMHS measures appropriate for use with preschool-aged children which were rated as having "excellent" psychometric accuracy, the *Pediatric Symptom Checklist* (PSC; Jellinek et al., 1988) and the *Strengths and Difficulties Questionnaire* (SDQ; Goodman & Goodman, 2009). Of these measures, only the SDQ was also identified as suitable for multi-informant use by parents and teachers. The current review builds on the work of Becker-Haimes et al. (2020) by specifically evaluating the effectiveness and acceptability of multi-informant screening measures in the preschool age range.

Multi-Informant Screening

Multi-informant report is a hallmark of developmental research and practice as it captures information about the child from multiple sources. Therefore, it is likely that UMHS for preschoolers will be particularly effective when multiple reporters are included (Scott et al., 2011). Such approaches incorporate the unique perspectives of the child's behaviour and wellbeing across settings such as the home and school or childcare by different caregivers (Anderson et al., 2019). Given the challenges of identifying normative versus concerning emotions and behaviour in the preschool-age bracket, secondary informants such as education or childcare providers are ideal as they may offer important and distinct insight into child MH in addition to parent perceptions (Feeney-Kettler et al., 2011; Smith, 2007). Educators are often familiar with children's social interactions, behaviour and wellbeing as they spend significant amounts of time observing children in naturalistic environments and in play with similar-aged peers. In Australia and the United States, the majority of 4–5 year old children are enrolled in preschool or early childhood education and care, attending between 15–21 h per week on average (Australian Bureau of Statistics, 2021; National Center for Education Statistics, 2016, 2018), again emphasising that educators are well-placed to conduct UMHS with an accessible cohort of children (DeLucia et al., 2022). However, there remains a need for children with MH difficulties to be systematically identified and referred by educators as part of an integrated model for child MH assessment (Casale & Reyes, 2023).

In contrast to primary care settings or formal MH services, preschools, pre-kindergarten, childcare, long daycare or early learning centres (henceforth referred to as 'preschools') are an ideal setting for UMHS since they are less encumbered by issues of access and stigma (Desta et al., 2017). Like schools for older children, preschools offer a unique opportunity to address the MH needs of children (der Embse & De Los Reyes, 2024). Easy access to preschools means that many parents turn to educators or preschool staff in the first instance to discuss MH concerns they have about their child(ren). In fact, research shows that parents frequently seek help from an informal source such as educators, friends or family, prior to seeking help from professional medical specialists, education or MH services (Pavuluri et al., 1996). Yet preschool educators have been subject to minimal research regarding their perspectives on child MH, and their role in identifying and supporting young children with socio-emotional or behavioural concerns (Croft et al., 2015). Instead, research to date on educator involvement in UMHS has focussed on primary and secondary school teachers. Research with older children has shown that over-reliance on single sources of information and decision-making based on non-scientific factors (i.e., factors which are not evidence-based such as small standardisation samples which are not representative of the target population) contributes to inequitable school MH decision-making (von der Embse & De Los Reyes, 2024). These findings are likely to be relevant for younger children as well, therefore emphasising the need for validated, multi-informant assessment.

Identifying emotional and behavioural concerns in children can be undertaken with varying degrees of accuracy by parents and other primary caregivers each with their own level of biases, knowledge and understanding of the child. Whilst preschool educators have a referent classroom group by which to compare the child, they may have little to no experience in the area of child MH (Stormont & Stebbins, 2005). Screening practices that rely exclusively on either parent only or educator only reports omit important information about child behaviour in different settings and are at risk of misidentifying children with MH concerns (Eklund & Dowdy, 2014; Stefan & Miclea, 2017). Thus, in order to identify children who may require support and additional MH services, parents and educators need clinically useful and effective screening measures that allow for multi-informant reports to help them accurately identify children at risk.

Clinical Utility and Effectiveness of Screening

A range of screening tools currently exist for identifying MH difficulties in young children; however, little is known about the multi-informant measures that exist, and the clinical utility and effectiveness of these measures. The clinical utility of these measures needs to be assessed to make value judgements of UMHS in this population group. However, "clinical utility" is a poorly defined construct in the UMHS literature and has been used to encompass an array of broad concepts

such as feasibility, practicability, acceptability, perceived utility and treatment effectiveness (Murphy et al., 1996; Proctor et al., 2011; Schubiner et al., 1994). This review defines "clinical utility" as whether the intended screening outcomes have perceived clinical value, that is whether they are helpful and accurate in identifying children at risk of MH difficulties (Gall et al., 2000; Humphrey & Wigelsworth, 2016; Schubiner et al., 1994). Psychometric properties of screening measures can thus be used to assess the accuracy or validity of measures (Humphrey & Wigelsworth, 2016).

When screening tools are inaccurate and thus lack clinical utility, they can incorrectly categorise children as at-risk when in fact they are not, an error that is known as false positives. Over-identification and risks of high false-positives in UMHS can result in wasted resources where children are unnecessarily triaged to receive intervention and/or stigmatisation of children who do not require help (Sawyer et al., 2014). Conversely, under-identification in the form of failing to identify children who are in fact at-risk, the error of false negatives, carries the risk of denying children access to treatment that may assist them. Over- and under-identification can be measured by the psychometric properties of predictive validity, including specificity and sensitivity. Reviewing measurement properties can be a time-consuming task requiring an understanding of psychometrics. This review's extraction of psychometric properties will, therefore, provide an accessible summary of the current literature.

In evaluating the clinical utility of multi-informant UMHS, the incremental validity of parent and educator report is of particular interest in considering the unique perspectives of each informant and what they add to an understanding of child MH. Adding educators' perspectives to parent-report has the potential to increase the validity of early identification, whilst also raising the possibility of discrepant ratings between informants (Croft et al., 2015; De Los Reyes et al., 2015). The presence of multi-informant discrepancy has been well-reported and in fact offers important, domain-relevant information specific to each informant's context e.g., home or school (De Los Reyes et al., 2023). However, the incremental, predictive validity of multi-informant approaches relative to the use of single informant report warrants further attention, especially in relation to young, preschool-age children.

Evaluating the effectiveness of an intervention can be assessed by its ability to achieve its intended outcomes (Andrews, 1999). More specifically, in the context of UMHS, effectiveness can be defined by the improvement of early identification of child MH concerns (Brinley et al., 2024). UMHS is thus considered effective if it improves the identification or diagnosis, referral, or treatment of child MH difficulties, which can be measured by rates of accurate identification of MH difficulties or risk, uptake of referral rates to MH services, or service uptake.

Whilst there are a number of previous reviews of screening by educators for Autism Spectrum Disorder in children, including preschool aged children (e.g., DeLucia et al., 2022), a recent systematic review examining the effectiveness and cost-effectiveness of school-based identification models including UMHS has found there is little research on the identification of MH difficulties in preschool-aged children (Anderson et al., 2019). As such there is a need to examine the literature regarding UMHS effectiveness in terms of service referral and uptake following identification.

Acceptability of Screening

The acceptability of UMHS can impact uptake and implementation of screening in clinical and community settings (Harrison et al., 2013). It has been argued that UMHS in young children is inappropriate and ineffective, may increase parental anxiety, stigmatise children with labels, and lead to the over-medication of children (Frances, 2012; Jureidini & Raven, 2012). These arguments made against UMHS have in some instances halted the implementation of UMHS in preschool-age children previously (Alexander & Mazza, 2015). As such it is essential to investigate the perceived acceptability of UMHS amongst users if it is to be adopted and implemented widely as an early intervention strategy.

The current review is primarily interested in users' perspectives of acceptability, that is, those of parents and educators. One element of acceptability comprises whether screening is considered appropriate, and/or perceived as helpful or useful. This may involve perceptions about whether the screening is wanted, needed or socially significant (Humphrey & Wigelsworth, 2016) or whether it is agreeable, palatable, or satisfactory (Proctor et al., 2011). Satisfaction can also encompass 'usability' of the tool (e.g., level of satisfaction with the length of screening, whether the language was easy to understand); whether informants felt comfortable or distressed by the screening; and whether informants would recommend screening to others, or if they would complete screening again in future (Brinley et al., 2024). Acceptability data, expressed as parental or educator attitudes towards screening, can be collected quantitatively or qualitatively. Importantly, acceptability encompasses characteristics that are associated with the likelihood of adoption of a screening measure (Glover & Albers, 2007; Kamphaus et al., 2007). This review will investigate the acceptability of UMHS broadly, and specific screening measures, as assessed by parents and educators.

No systematic review to date has evaluated multi-informant UMHS measures by parents and educators of preschool-aged children. As such, there is a need to identify what measures exist for children in this age range, examine the

clinical utility and effectiveness of these measures, and also to examine the acceptability of UMHS more broadly.

The Current Study

The aim of this systematic review is to examine multi-informant UMHS for preschool-aged children (3–5 years) by both parents and educators. The review will answer the following questions:

1. What are the existing MH screening measures utilised by both educators and parents of preschool-aged children and what is the clinical utility of these measures, that is, what is the predictive and incremental validity of these measures?
2. What is the effectiveness of UMHS?
3. What is the acceptability of utilising UMHS amongst educators and parents of preschool-aged children?

Methods

The protocol for this systematic review was registered with the International prospective register of systematic reviews, PROSPERO, on 22 December, 2022 (Registration number CRD42022383426).

Eligibility Criteria

Peer-reviewed studies examining UMHS for child mental health or socio-emotional and behavioural wellbeing in preschool children aged 3 years 0 months—5 years 11 months were included. Included studies reported universal screening whereby all children within a population-based sample undertook the systematic assessment of MH or socio-emotional or behavioural wellbeing and ratings were provided by both parents and educators.

Studies were excluded if they focussed specifically on Autism Spectrum Disorder or were not published in English. Complete inclusion and exclusion criteria are presented in Table 1.

Information Sources

Seven databases were searched for relevant literature: CINAHL, Embase, ERIC, Medline, PsycINFO, Scopus and Web of Science. All databases were searched via Ovid aside from CINAHL, Scopus and Web of Science.

Search Strategy

The first author developed and translated the search strategy according to the relevant subject headings and functionality of each database in consultation with two authors (LT and MD) and a research librarian. The strategy was trialled in October and December 2022. Databases were searched in May 2023 using the following search strategy:

1. (exp mass screening/ or mass screening*.mp. or universal screening*.mp. or early intervention*.mp. or early identification*.mp. or screening*.mp. or exp needs assessment/ or needs assessment*.mp OR (mass adj3 screening*) OR (universal adj3 screening*)).
AND
2. (exp mental health/ OR mental health.mp OR exp mental disorder/ OR wellbeing.mp OR wellbeing.mp OR well-being.mp OR ((mental or psychological or behavioral or socio-emotional or social?emotional or socioemotional) adj2 (disorder* or difficult* or health* or problem* or wellbeing or well-being)).mp).
AND
3. (preschool student*.mp OR exp Child, Preschool/ OR preschool*.mp OR pre-kindergarten*.mp OR prekindergarten* OR early childhood*.mp OR early learning*.mp).
AND
4. (exp school teachers/ OR teacher*.mp OR educator*.mp OR child care worker*.mp OR day care worker*.mp OR daycare worker*.mp OR childcare worker*.mp OR "prekindergarten teacher*" OR "pre-kindergarten teacher*").

Results were limited to peer reviewed journals.

To ensure literature saturation, we scanned the reference lists of included studies identified through the search and conducted manual handsearching.

Study Selection

Article citations from the search results were uploaded to Covidence (Veritas Health Innovation). After duplicates were removed, the first author screened all papers by reviewing the title and abstract of papers and a team of five authors (SB, TC, RM, LT and AT) independently screened papers against criteria. A calibration exercise was undertaken to pilot and refine the screening questions. Studies not meeting criteria were excluded.

Full texts for the remaining studies were obtained and examined against criteria. Studies not meeting the criteria were excluded using a hierarchy of exclusion criteria as outlined in Table 1. All papers were screened by two reviewers to reduce selection bias. The first author screened all papers and the same team of five authors

Table 1 Inclusion and Exclusion Criteria

Inclusion criteria:	Exclusion criteria (in ranked order):
	Duplicate studies
Studies published in English	Studies not published in English
Studies published in peer-reviewed journals	Studies not published in peer-reviewed journals; grey or unpublished literature (dissertations, theses)
Full text available	No full text available
Empirical data presented	No empirical data presented (Systematic review/meta-analysis; Case study; Narrative review; Conference proceedings; Commentary; Protocol paper)
Studies focussed on individuals in the age range 3–5 years (mean age ≤ 3 years 0 months and ≥ 5 years 11 months, or more than half the sample within the age range)	Studies focussed on individuals outside of the age range 3–5 years (mean age ≤ 3 years 0 months and ≥ 5 years 11 months, or more than half the sample outside the age range). This includes excluding any studies in which:
	a. No age details for child samples are reported and individuals are not attending preschool or similar early childhood learning or childcare;
	b. No age details for child samples are reported and children are enrolled in kindergarten, elementary grade or above (e.g., grade one; secondary school, etc.);
	c. Sample (or sub-sample analyses) do not report mean and age range includes children under 2 years and 0 months or children over 6 years and 11 months
	Screening with wrong primary respondent such as Medical Professional (GP, Paediatrician, Nurse, etc.), Allied Health (OT, Social Work, etc.), Psychologist, or Child report
Screening with parent and educator report	Screening with parent report only or educator report only
Studies in which the primary focus is screening	Studies in which the primary focus is not screening (e.g., prevalence surveys)
Studies in which the screening is universally implemented	Studies in which the screening is not universally implemented (i.e., screening is implemented with targeted populations)
Screening with a primary focus on child mental health or socio-emotional and behavioural wellbeing	Screening with a primary focus other than child mental health or socio-emotional and behavioural wellbeing (e.g., parenting, parental mental health, physical health, academic performance)
	Studies which do not report predictive validity, incremental validity, effectiveness or acceptability data for screening measures
	Studies which focus on screening for Autism Spectrum Disorder

independently screened a selection of papers against criteria. At either stage of screening, a third reviewer was consulted when consensus could not be reached by prior discussion.

The study selection process is outlined in Fig. 1.

Data Collection

Data for the included studies were extracted in excel using a codebook developed by the authors. Six authors (SB, TC, RM, JN, LT and AT) independently conducted data extraction. Extraction into the codebook was trialled twice using two included studies. Codebook headings are included in Supplementary file 1. Our data collection only utilised published data and we did not contact study authors for additional information.

Study Quality Assessment

All included studies were critically appraised for risk of bias using the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018), an evaluation tool designed for systematic reviews that include mixed studies i.e., qualitative, quantitative or mixed methods studies. The MMAT comprises two stages. First, two initial screening questions confirm that the study is an empirical study and appropriate to be appraised by the MMAT. Second, for each included study, the appropriate methodological category (e.g., quantitative descriptive, quantitative non-randomised, qualitative) is selected before answering corresponding questions. A calibration exercise was undertaken to ensure all reviewers understood how to assess studies using the MMAT and resulted in an acceptable level of agreement between raters.

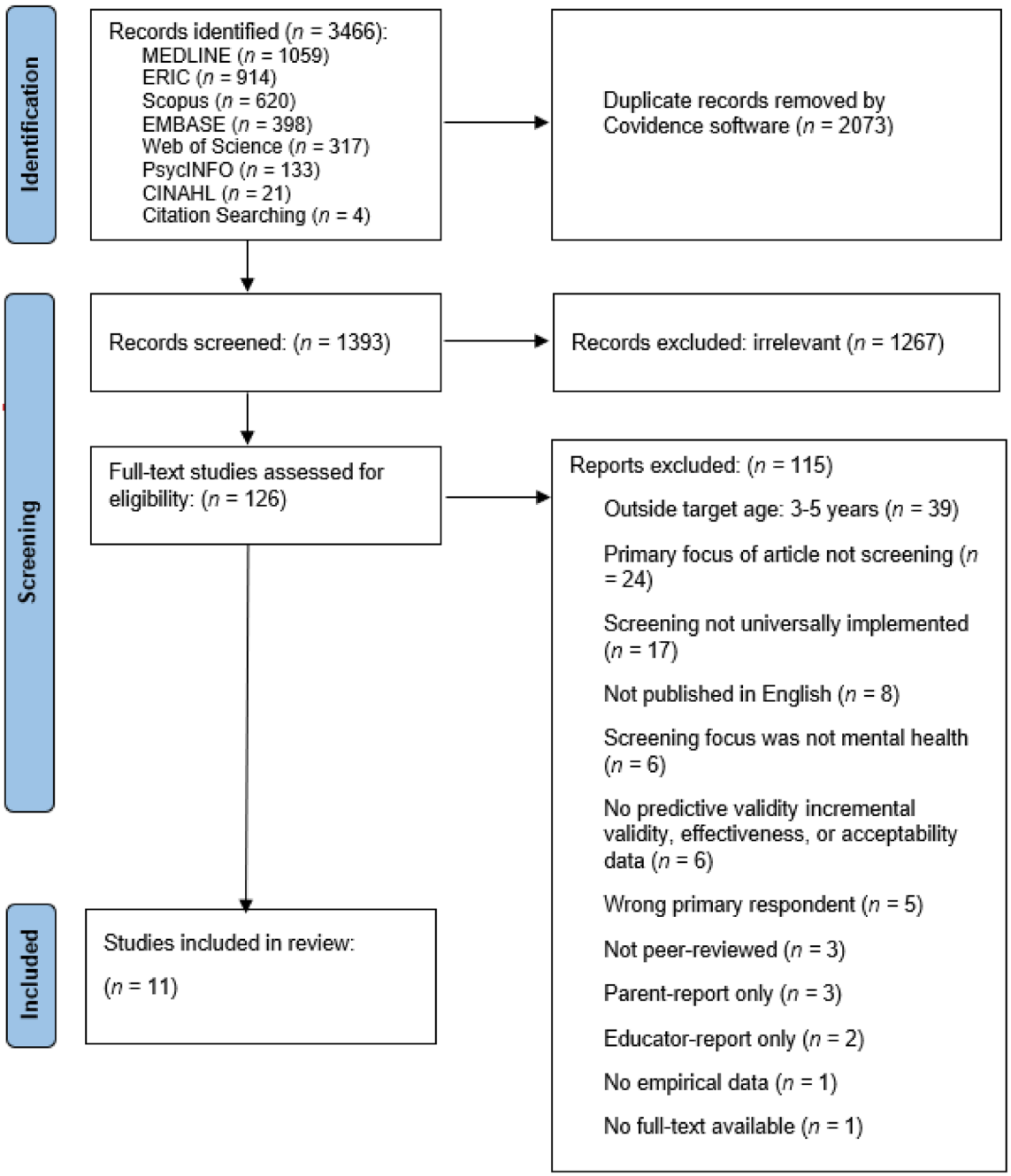


Fig. 1 PRISMA Flow Diagram for Identification of Studies

Results

Study Selection

Database searches yielded 3,466 results. After duplicates were removed, 1,393 were screened including 126 that underwent full-text review. The final number of studies included was 11. The study selection process is illustrated in Fig. 1.

Study Characteristics

All studies were published within the past 10 years, aside from two studies published in 2007 (Barbarin) and 2011 (Feeney-Kettler et al.). Of the 11 studies included in the review, over half of the studies were conducted in the United States ($n = 7$). Studies were also conducted in Japan ($n = 1$), the Netherlands ($n = 1$), Romania ($n = 1$),

and Spain ($n=1$). There were no studies conducted in low and middle income countries.

Mental health domains screened in the included studies were mainly broad social, emotional, and behavioural wellbeing ($n=10$), and ADHD ($n=1$). All children were community samples recruited from either preschools or kindergartens ($n=8$), community health clinics ($n=2$) or random census sampling ($n=1$). Of the studies reporting demographic details about parents, 70% of recruited samples were predominantly female. Of the studies that reported child gender ($n=6$), all recruited samples were gender balanced i.e., between 45–55% female/male. No study reported non-binary gender for children. The socio-economic status and cultural diversity of samples varied across studies. A summary of study characteristics is provided in Table 2.

Results of Individual Studies

Overview of Measures

The research studies examined 10 measures. As is expected for universal screening measures of community populations, almost all measures assessed broad constructs of internalising and externalising difficulties, including the *Attention, Behaviour, Language, Emotions* (ABLE; Barbarin, 2007); *Behaviour Assessment System for Children* (BASC; Moore et al., 2022a, 2022b); *Behavior Assessment System for Children—Behavioral and Emotional Screening System*, (BASC-2—BESS; Kettler et al., 2017); *Emotion Competence Screening* (ECS; Stefan & Miclea, 2017); *Impairment Rating Scale* (IRS; Girio-Herrera et al., 2015); *Parents' Evaluation of Developmental Status* (PEDS; Doove et al., 2019); *Pediatric Symptom Checklist-17* (PSC-17; Moore et al., 2022a); *Preschool Behavior Screening System* (PBSS; Feeney-Kettler et al., 2011, 2019; Kettler et al., 2017); *and Strengths and Difficulties Questionnaire* (SDQ; Ezpeleta et al., 2013). One measure focussed on attention or hyperactivity, *ADHD-Rating Scale-IV* (Takayanagi et al., 2016). Table 3 details the measures used in the studies. Some measures also included prosocial or adaptive behaviour (ECS, IRS, PBSS, SDQ); and the PEDS measured developmental domains in addition to social, emotional and behavioural constructs (Doove et al., 2019). The number of items in each screening measure varied widely, particularly as some measures were multiphasic and included additional items dependent on previous responses (e.g., for each issue that a respondent indicated concern, follow-up questions were then asked about the level of severity and impact on the child's life).

Most studies reported inter-rater reliability between parent and educator informants using Pearson's correlations. Two studies reported moderate cross-informant agreement between parent and educator ratings (Barbarin, 2007; Feeney-Kettler et al., 2011). Barbarin (2007) reported that parents and educators agreed 77% of the time about children who did not have difficulties. One study presented chi-square analyses of cross-informant agreement, reporting a statistically significant association between parental concerns and professional caregivers' concerns about child wellbeing and development at baseline and 10-month follow-up ($\chi^2=34.8$; df = 1, $p<0.001$ and $\chi^2=8.1$; df = 1, $p=0.004$) (Doove et al., 2019).

The majority of studies ($n=7$) that reported internal consistency for parent-report measures reported high reliability (Cronbach's alphas above 0.8) (Ezpeleta et al., 2013; Feeney-Kettler et al., 2011, 2019; Kettler et al., 2017; Moore et al., 2022a; Stefan & Miclea, 2017; Takayanagi et al., 2016). Most studies ($n=7$) that reported internal consistency for educator-report measures reported high or very high reliability (Cronbach's alphas above 0.8) (Ezpeleta et al., 2013; Feeney-Kettler et al., 2011, 2019; Kettler et al., 2017; Moore et al., 2022a; Stefan & Miclea, 2017; Takayanagi et al., 2016). Three studies did not report internal consistency for either parent or educator report (Barbarin, 2007; Girio-Herrera et al., 2015; Moore et al., 2022b).

Few studies reported test–retest reliability (Doove et al., 2019; Feeney-Kettler et al., 2011; Stefan & Miclea, 2017). Of the three studies that did, all reported good to excellent reliability for parent report and three studies reported good to excellent reliability for educator report. See Supplementary file 1 for further details regarding screening measures' inter-rater reliability, internal consistency, and test–retest reliability.

Clinical Utility

To assess clinical utility, the predictive and incremental validity of measures were extracted. Studies utilised a range of criterion measures to test the predictive validity of their screening measures. Criterion measures varied widely with only one study utilising gold standard diagnostic interviews (Ezpeleta et al., 2013) or comprehensive assessments ($n=6$) such as the Child Behavior Checklist (CBCL;Achenbach, 1999) or Caregiver-Teacher Rating Form (C-TRF; Achenbach & Rescorla, 2000), whilst others utilised other brief screening measures such as the PSC-17 or SDQ ($n=5$) or a combination of measures. See Table 3 for details.

Seven out of the eight studies examining predictive validity reported acceptable or good sensitivity or specificity for parent ratings. Kettler et al. (2017) reported the strongest predictive validity for children in the clinical range: sensitivity 90%, specificity 90%, Positive Predictive Value (PPV) 42% and Negative Predictive Value (NPV) 99%. This study tested the PBSS against a comprehensive range of criterion measures including the BASC-2—BESS, C-TRF and CBCL 1.5–5, which included both parent and educator reports.

Table 2 Study Characteristics

Author (Year)	Country	Study Design	Population & Setting	Study Aims	Child Characteristics		
					Sample size (n)	Gender* (female, male)	Age range (M, SD)
Barbarin (2007)	USA	Cross-sectional analytical study	Study 1: Pre-K children in 40 randomly selected classrooms in 6 participating states Study 2: Children enrolled in the Head Start program or early childhood programs receiving partial state financial support	To determine the nature and prevalence of socio-emotional concerns parents and educators have about preschool children, the degree to which parents and educators agree on their concerns, and the psychometric properties of the ABLE	Study 1: 415 Study 2: 5,577	Study 1: NR Study 2: NR	Study 1: 4 yrs (NR) Study 2: 3–4 yrs (NR)
Doove et al., (2019)	Netherlands	Prospective observational study	Children aged 3 years at study entrance as part of the Monitoring Outcome Measurements (MOM) child development study	To define psychometric properties of the Dutch PEDS and three VAS about 'parenting', 'child behaviour' and 'child competence' at the age of 3 and 4 years	346	52%, 48%	3 years (3.0; 0.2)
Ezpeleta et al., (2013)	Spain	Cross-sectional analytical study	A random sample of 3-year-olds from the census of preschoolers in Barcelona	To evaluate the psychometric properties of the SDQ in sample of Spanish preschool children	Phase 1: 1,341 Phase 2: 622	Phase 1: NR, 50.9% Phase 2: NR, 50%	Phase 1: 3 years (3.0; 0.18) Phase 2: 3 years (3.0, 0.16)
Feeney-Kettler et al., (2011)	USA	Quantitative Descriptive	Children in $n = 22$ preschools and childcare facilities in Southern California and Tennessee	To streamline the PBSS Phase 2 to create separate parent & educator versions, and to evaluate psychometric properties of the PBSS as a cost-efficient universal screening tool	NA	49%, 51%	3–5 years (NR)
Feeney-Kettler et al., (2019)	USA	Quantitative Descriptive	Preschool Children recruited from daycares and preschools (number not reported) in a large southern city and its surrounding suburbs in the US	To evaluate the multiple-gate PBSS for identifying children's social, emotional, and behavioral difficulties	122	48%, 44%	3–5 years (NR)

Table 2 (continued)

Author (Year)	Country	Study Design	Population & Setting	Study Aims	Child Characteristics		
					Sample size (n)	Gender* (female, male)	Age range (M, SD)
Girio-Herrera et al., (2015)	USA	Cross-sectional analytical study	Kindergarteners at $n=18$ elementary schools (Study 1) $n=5$ elementary schools (Study 2)	To examine the IRS as a screening tool for detecting kindergarten children who are at risk for social, emotional, academic, and behavioral difficulties	NA	Study 1: NR, 46.8% Study 2: NR, 50.8%	5 years 1 month–5 years 11 months Study 1: (5.48, 0.32) Study 2: (5.61, 0.46)
Kettler et al., (2017)	USA	Cross-sectional analytical study	Preschool children taken from a convenience sample of $n=33$ preschools in a heavily populated Northeastern US state	To determine the internal consistency, cross-informant agreement of a multi-gate screener and the concurrent relations among scores from single-gate and multi-gate screeners	105	NR, 60%	3–5 years (NR)
Moore et al. (2022a)	USA	Cross-sectional analytic and longitudinal cohort study	Preschool children from $n=5$ state-funded preschools from a Title 1 (low income supported) school district	To examine the use of informants at the preschool level when universally screening for behavioral and emotional risk	535	48.3%, 44%	3–5 years (4.32, NR)
Moore et al. (2022b)	USA	Quantitative Descriptive	Preschool, Kindergarten, and first-grade children from $n=5$ state-funded preschools housed within public elementary schools	A preliminary examination of educators' and parents' beliefs related to their participation in Universal Mental Health Screening	NA	NR, NR	NR (NR)
Stefan et al., (2017)	Romania	Cross-sectional analytical study	Preschool children from $n=3$ preschools within an urban area of Cluj-Napoca	To evaluate the reliability and validity of the ECS and SCS parent and educator versions	Sample 1: 107 Sample 2: 73	Sample 1: 48.6%, 51.4% Sample 2: 54.8%, 45.2%	Sample 1: 2 years, 4 months–4 years (43.84 months, NR) Sample 2: 4–5 years (4.33, NR)
Takayanagi et al., (2016)	Japan	Quantitative Descriptive	Five-year-old children from a local community health check-up run through municipal health centres in Aomori prefecture	To verify the psychometric properties of the ADHD-RS in identifying preschool children with ADHD using DSM-5 criteria at a community health check-up for five-year-olds	838	46.1%, 53.9%	4 years. 10 months–5 years, 10 months (4.83, 0.28)

Kettler et al. (2017) also reported the strongest predictive validity for educator ratings for children in the clinical range: sensitivity 97%, specificity 91%, PPV 45% and NPV 100%. When compared to the criterion measure, educators were slightly more accurate raters than parents.

The sensitivity of the SDQ was poor for both parent and educator reports, across all subscales aside from the prosocial subscale (Ezpeleta et al., 2013). One study reported combined predictive validity for parent and an educator reports incorporating multi-informant reports to identify children at risk of internalising and externalising difficulties, with the PBSS showing combined sensitivity 98% and specificity 28% (Feeney-Kettler et al., 2011). Three studies did not report predictive validity (Feeney-Kettler et al., 2019; Moore et al., 2022a, 2022b).

The majority of studies ($n = 8$) reported Area Under Receiver Operating Characteristic Curve (AUC) analyses as an indicator of screening accuracy. Criterion measures again varied widely as did the level of analyses conducted (e.g., informant report, subscale, child gender, etc.). All studies reporting AUC analyses for parent report ($n = 7$), reported fair (above 0.70) to excellent (0.90–1.00) AUCs. Five studies reporting AUC analyses for educator report, reported very good to excellent AUCs (0.80–1.00) (Doove et al., 2019; Feeney-Kettler et al., 2011, 2019; Kettler et al., 2017; Stefan & Miclea, 2017). Two studies reported AUCs that were below acceptable levels for educator report (Ezpeleta et al., 2013; Takayanagi et al., 2016). Finally, only one study (Girio-Herrera et al., 2015) tested cross-informant ratings by parent-reported Impairment Rating System (IRS) identifying educator-rated BASC 2 at-risk status, and educator IRS identifying parent-based BASC 2 at-risk status—both resulting in poor AUCs.

Incremental Validity

Only one study reported the incremental validity of parent and educator reports in terms of child social-emotional outcomes (Moore et al., 2022a). Using hierarchical regression modelling, this study reported that educator-report of social-emotional difficulties using the BASC-3 BESS *Behavioural and Emotional Risk Index* (BERI) was significantly associated with kindergarten social-emotional readiness ($\beta = -0.46$), however including parent-report BESS BERI did not significantly improve prediction of kindergarten social-emotional readiness. However, when parent-report was entered in the first block, parent-report was significantly associated with kindergarten social-emotional readiness ($\beta = -0.19$) and when educator ratings were added in the second block, educator-report was also significantly associated with kindergarten social-emotional readiness ($\beta = -0.44$) although parent ratings

Table 2 (continued)

*no studies reported non-binary child gender. *ABLE* Attention Behaviour Language Emotions, *ADHD* Attention-Deficit/Hyperactivity Disorder, *ADHD-RS* Attention-Deficit/Hyperactivity Disorder Rating Scale, *DSM-5* Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, *ECS* Emotion Competence Screening, *IRS* Impairment Rating Scale, *NR* not reported, *PBSS* Preschool Behavior Screening System, *PEDS* Parents' Evaluation of Developmental Status, *SCS* Social Competence Screening, *US* United States, *VAS* Visual Analogue Scales

Table 3 Screening Measure Properties

Author (Year)	Sample N (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
Barbarin (2007)	Study 1: 238, 415 Study 2: NR, NR	ABLE (Stage 1: 10, Stage 2: 40)	Attention, Behaviour, Language, Emotions	Study 1 TCRS; BPI; ORCB; PPVT; OWLS Study 2: N/A	**Educator Pre-K**: 0.42 **K:** 0.65 **Parents:** 0.59 (end of Pre-K) 0.60 (end of K)	**Parent:** good .86 (end of pre-K), .84 (end of K)	NR NR	NR NR
Doove et al., (2019)	294, 329	VAS (3) PEDS– Dutch Version (10)	School Readiness/ Social Participation	CBCL SDQ	**Educator:** Child competence VAS 82.8 PEDS: 96.6 **Parent:** Parenting VAS: 90.9 Child behaviour VAS: 71.4 PEDS: 80.6	**Educator** Child competence VAS 82.8 PEDS: 83.1 **Parent:** Parenting VAS: 78.4 Child behaviour VAS 80.8 PEDS: 80.2	NR NR	**Educator** Child competence VAS 97.6 PEDS: 99.5 **Parent** Parenting VAS: 99.2 Child behaviour VAS 95.9 PEDS: 98.9
Ezpeleta et al., (2013)	**Phase 1** N/A, 1,341 **Phase 2** 94, 622	SDQ (25)	Behaviour Difficulties	CBCL (1.5—5 years); CGAS; DICA-PPYC; SDQ (Spanish and Catalan versions)	**Educator** Screening cutoff: 72.1 Borderline: 38.2 Abnormal: 18.5 **Parent** Screening cutoff: 74.2 Borderline: 52.1 Abnormal: 29.9	**Educator** Screening cutoff: 44.3 Borderline: 78.3 Abnormal: 89.2 **Parent** Screening cutoff: 62.1 Borderline: 83.4 Abnormal: 95.8	NR NR	NR NR
Feeney-Kettler et al., (2011)	112 113	PBSS (46)	Internalising Externalising Prosocial Behaviour	BASC-2	**Educator Tot** = .94 Int = .96 Ext = 1.00 **Combined Ed/P** Tot = .98 Int = .89 Ext = .98 **Parent:** Tot = .96; Int = .80 Ext = .90	**Educator Tot** = .51 Int = .59 Ext = .71 **Combined Ed/P** Tot = .28 Int = .42 Ext = .48 **Parent:** Tot = .49 Int = .61 Ext = .59	**Educator Tot** = .49 Int = .44 Ext = .54 **Combined Ed/P** Tot = .57 Int = .62 Ext = .57 **Parent:** Tot = .39 Int = .57 Ext = .51	**Educator Tot** = .95 Int = .98 Ext = 1.00 **Combined Ed/P** Tot = .93 Int = .77 Ext = .97 **Parent:** Tot = .96 Int = .82 Ext = .93

Table 3 (continued)

Author (Year)	Sample N (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
Feeney-Kettler et al., (2019)	122 122	PBSS (46)	Internalising, Externalising, Prosocial Behaviour	BASC-2	NA NA	NA NA	NA NA	NA NA
Girio-Herrera et al., (2015)	**Study 1:** 56 12 **Study 2:** 568 273	IRS (31)	**Study 1**: Academic, Social, Behaviour, Family **Study 2:** Academic, Social, Behaviour, Family	**Study 1:** BASC-2 **Study 2:** BESS	**Study 1:** **Educator** 0.57 **Parent** 0.14 **Study 2:** **Educator** 0.72 **Parent** 0.17	**Study 1:** **Educator** 0.92 **Parent** 0.98 **Study 2:** **Educator** 0.95 **Parent** 0.95	**Study 1:** **Educator** 0.65 **Parent** 0.98 **Study 2:** **Educator** 0.65 **Parent** 0.95	**Study 1:** **Educator** 0.89 **Parent** 0.74 **Study 2:** **Educator** 0.97 **Parent** 0.90
Kettler et al., (2017)	105 105	PBSS (46)	Social, Emotional, and Behaviour Difficulties	BASC-2, C-TRF, CBCL	**P1 + P2** **Educator** ASEBA TP Borderline = 0.76, Clinical = 0.97 **Parent** ASEBA TP Borderline = 0.66, Clinical = 0.90	**P1 + P2** **Educator** ASEBA TP Borderline = 0.96, Clinical = 0.91 **Parent** ASEBA TP Borderline = 0.91, Clinical = 0.90	**P1/P2** **Educator** ASEBA TP Borderline = 0.80, Clinical = 0.45 **Parent** ASEBA TP Borderline = 0.91, Clinical = 0.90	**P1/P2** **Educator** ASEBA TP Borderline = 0.95, Clinical = 1.00 **Parent** ASEBA TP Borderline = 0.95, Clinical = 0.99
Moore et al. (2022a)	14 535	BASC-3 BESS (86); PSC-17 (17)	Behavioural and Emotional Risk	BASC-3 BESS	NR NR	NR NR	NR NR	NR NR
Moore et al. (2022b)	40 330	BASC-3 BESS (86)	Behavioural and Emotional Risk	PSC-17	NA NA	NA NA	NA NA	NA NA

Table 3 (continued)

Author (Year)	Sample N (Educator, Parent)	Screening measure (no. of items)	Mental health domain	Predictive validity				
				Criterion measure	Sensitivity (Educator, Parent)	Specificity (Educator, Parent)	PPV (Educator, Parent)	NPV (Educator, Parent)
Stefan et al., (2017)	NR 180	ECS (30)	Social and Emotional Competencies	SRSS	**Educator** ECS-T to C-TRF: Sensitivity indices between 70.00—75.00 (internalising), and 75.00—83.33 (externalising) SCS- T to C-TRF: 75.00 to 77.78 (internalising); and 71.43- 75.00 (externalising) **Parent** ECS-P to CBCL: 80.00—84.62 (internalising); and 69.23—71.43 (externalising) SCS- P to CBCL: 76.92—80.00 (internalising); and 71.43—75.00 (externalising)	**Educator** ECS-T to C-TRF: 92.78–93.85 (internalising). and 92.21—94.74 (externalising) SCS-T to C-TRF: 91.84–94.20 (internalising), and 90.79—91.58 (externalising) **Parent** ECS-P to CBCL: 92.65—93.62 (internalising), and 92.42%—92.47 (externalising) SCS- P to CBCL: 89.36—92.65 (internalising), and 88.42%—93.94 (externalising)	**Educator** ECS-T to C-TRF: 50.00–60.00 (internalising); 45.45–64.29 (externalising) SCS-T to C-TRF: 42.86–46.67 (internalising); 41.67–52.94 (externalising) **Parent** ECS-P to CBCL: 44.44–64.71; (internalising); 50.00–56.25 (externalising) SCS-P to CBCL: 44.44–50.00 (internalising); 45.00–55.56 (externalising)	**Educator** ECS-T to C-TRF: 96.53–96.77 (internalising); 96.77–98.61 (externalising) SCS-T to C-TRF: 97.83–98.48 (internalising); 96.67–97.18 (externalising) **Parent** ECS-P to CBCL: 97.78–98.44 (internalising); 95.56–96.83 (externalising) SCS-P to CBCL: 96.55–98.44 (internalising); and 96.55–96.88 (externalising)
Takayanagi et al., (2016)	NR 838	ADHD-Rating Scale-IV Parent (18); Educator (18)	ADHD	SDQ	**Educator** Above 90th percentile 30.23 **Parent** Above 90th percentile 89.13	**Educator** 90.92 **Parent** 94.07	**Educator** 16.05 **Parent** 46.59	**Educator** 95.78 **Parent** 99.33

Data reported for recommended cutoff scores or best tradeoff between sensitivity and specificity. *ABLE* Attention Behaviour Language Emotions, *ADHD* Attention-Deficit/Hyperactivity Disorder; *ASEBA* Achenbach System for Empirically Based Assessment, *BASC-2/3* Behaviour Assessment System for Children Second/Third edition, *BESS* Behavioral and Emotional Screening System, *BPI* Behaviour Problem Index, *CBCL* Child Behavior Checklist, *CGAS* Children's Global Assessment Scale, C-TRF Caregiver-Teacher Rating Form, *DICA-PPYC* Diagnostic Interview for Children and Adolescents for Parents of Preschool And Young Children, *ECS-P/T* Emotion Competence Screening Parent/Teacher Form, *Ext* externalizing, *K* Kindergarten, *Int* internalizing, *IRS* Impairment Rating Scale, *NR* not reported, *ORCB* Observer ratings of Child Behaviour, *OWLS* Oral and Written Language Scale, *PBSS* Preschool Behavior Screening System, *PEDS* Parents' Evaluation of Developmental Status, *PPVT* Peabody Picture Vocabulary Test, *PSC-17* Pediatric Symptom Checklist-17, *SCS-P/T* Social Competence Screening Parent/Teacher Form, *SDQ* Strengths and Difficulties Questionnaire, *SRSS* Social Skills Rating System, *T* total problems, *TCRS* Teacher Child Rating Scale, *VAS* Visual Analogue Scales

in the second block were no longer significantly related to readiness ($\beta = -0.07$, $p = 0.26$).

In addition, Moore et al. (2022a) reported incremental validity using a second screening measure, the PSC-17. Educator-reported social-emotional difficulties using the PSC-17 was significantly associated with kindergarten social-emotional readiness ($\beta = -0.43$) and parent report was not associated with kindergarten social-emotional readiness nor did inclusion result in significant improvements in variance in the model ($\beta = 0.05$, $R^2 = 0$). When the parent report was entered in the first block, it was not statistically significant. However, when educator ratings were added to subsequent blocks, educator ratings were significantly associated with kindergarten social-emotional readiness ($\beta = -0.44$) and parent ratings were no longer significantly related to social-emotional readiness. It is important to note that kindergarten social-emotional readiness is a criterion variable that references child behaviour in the education context only and therefore criterion contamination (whereby differences in parent and educator ratings may have arisen due to educators' influence over the criterion variable, in this case social-emotional readiness) may be a factor in these findings which seemingly favour educator report. As Moore et al. (2022a) conclude, both parents and educators offer valuable information when screening of a child's MH risk. However, educator report may be more informative than a parent report for predicting kindergarten social-emotional readiness, which we note is an education-specific measure of MH. It is also important to note that this study did not report other psychometrics of interest such as predictive validity more generally.

Effectiveness

Only one study reported the effectiveness of the screening measure in terms of referral uptake and longitudinal outcomes for children after screening preschool children using the Attention, Behavior, Language, and Emotions (ABLE) (Barbarin, 2007). However, this study only focussed on parent reports, and did not include the effectiveness of educator reports or multi-informant data. Moreover, this study reported effectiveness in terms of referral for Individualised Education Plans (IEP), the same outcome used to test predictive validity of the ABLE. After conducting screening for preschool children, 13.1% of children who were identified as having serious concerns through parent-reported ABLE were referred for IEP by the end of kindergarten. This was compared to 3.7% of children for whom no concerns were identified. The study did not report if this was a significant difference. The remaining studies did not report effectiveness data about screening measures.

Acceptability of UMHS

Only three studies examined the acceptability of screening measures or UMHS in general amongst parents and educators. Two studies investigating the acceptability of the PBSS measure found that 88% of parents in one sample ($n = 107$) and 87% of parents in another sample ($n = 91$) reported that the PBSS could provide useful information about their child, and 91% ($n = 111$) and 94% ($n = 99$) found it clearly written (Feeney-Kettler et al., 2019; Kettler et al., 2017). Amongst educators, 88% ($n = 107$) and 71% ($n = 74$) indicated that the PBSS could provide useful information about preschool children, and 89% ($n = 109$) and 88% ($n = 83$) found it clearly written (Feeney-Kettler et al., 2019; Kettler et al., 2017).

The initial study reported that just over half of educators (55%, $n = 58$) would be likely to use the PBSS to screen all their students for social and emotional wellbeing and some educators indicated that they would select students to screen rather than implement screening universally (Kettler et al., 2017). The second study reported that 71% of educators ($n = 87$) indicated that they would be likely to use the PBSS to screen all students (Feeney-Kettler et al., 2019). Neither study reported the number of items in the acceptability or evaluation measure.

A third study measured acceptability of UMHS more generally using four items encompassing constructs of importance, usefulness, willingness and appropriateness (Moore et al., 2022b). On average, educators and parents agreed or strongly agreed that it is important for schools to ask questions about child MH, screening is useful for identifying at-risk children, and MH development should be addressed in school (Moore et al., 2022b). Only 0.3–1.8% ($n = 6$) of parents and 0–10.0% ($n = 4$) of educators disagreed or strongly disagreed with the acceptability of UMHS (Moore et al., 2022b).

Risk of Bias in Studies

All studies were assessed using the MMAT (Hong et al., 2018). All studies were considered moderate to high quality. Quantitative descriptive studies tended to rate poorly for the risk of non-response bias or explaining non-responders and low response rates. Non-randomised trials frequently did not adequately account for confounders in the design or analysis. There were no randomised trials identified.

Discussion

The aim of this review was to systematically examine research conducted on multi-informant UMHS for preschool-aged children by both parents and educators. We identified which measures have been utilised in the preschool

population and examined their clinical utility in terms of predictive and incremental validity and effectiveness. The review also examined the acceptability of these measures amongst parents and educators. The recency of included studies suggests that this is an emerging area of research, conducted mainly in the United States and in wealthy, industrialised nations. Identified screening measures were highly varied and for some aspects examined, such as incremental validity, effectiveness and acceptability, there was a paucity of research making definitive conclusions difficult. However, there were 11 included studies and 10 measures identified for multi-informant UMHS in the preschool period, with promising findings for some measures suggesting that UMHS can accurately identify child MH concerns using parent and educator reports.

In order to assess the clinical utility of screening, this review examined the predictive and incremental validity of MH screening measures. Several screening measures demonstrated strong predictive validity for identifying children at-risk of MH difficulties. However, as is common for predictive validity of screening measures, several studies reported either high specificity or high sensitivity due to the tendency of sensitivity and specificity to be inversely related, and the challenge of developing a measure with maximal precision. Interestingly, the research showed that more educator-report measures had strong predictive validity (Girio-Herrera et al., 2015; Kettler et al., 2017; Stefan & Miclea, 2017) compared to parent-report (Kettler et al., 2017; Takayanagi et al., 2016). Using the benchmark of at least 70% sensitivity, 70% specificity and 50% PPV for developmental screening measures (Aylward, 1997), the ECS, IRS and PBSS met these standards for educator ratings (Girio-Herrera et al., 2015; Kettler et al., 2017; Stefan & Miclea, 2017). On the other hand, only the PBSS met this benchmark for parent-report (Kettler et al., 2017). However, results for this measure were mixed as earlier studies found that the combined, multi-informant predictive validity for the PBSS did not meet standards (Feeney-Kettler et al., 2011).

Surprisingly, the preschool version of the widely used SDQ demonstrated poor predictive validity with low sensitivity across both parent and educator reports and borderline acceptable to good specificity for parents and educators (Ezpeleta et al., 2013). Only the hyperactivity subscale had acceptable sensitivity and specificity at the recommended cut points and the study reported poor educator-report screening accuracy in terms of AUC analysis. Other comprehensive reviews of the SDQ as a screening measure amongst preschool populations have found a number of problems with its predictive validity, internal consistency of its subscales, test–retest reliability and criterion (concurrent) validity and have cautioned its use within non-clinical populations (Kersten et al., 2016; Lavigne et al., 2016). The preschool version of the SDQ may not have strong clinical utility or other psychometric properties as it has not been specifically developed for preschool samples, but further research is required to investigate this.

Only one included study reported on the incremental validity of adding educators' ratings to parent ratings, finding a significant improvement in the identification of socio-emotional outcomes over time (Moore et al., 2022a). Contrary to previous research with older children, this study found that parent ratings did not significantly improve the prediction of outcomes. Previous findings with older children reported that parent ratings of preschool children added significant variance to child outcomes when considered in addition to later kindergarten teacher ratings (Owens et al., 2015) and a large, longitudinal study of children aged 4–11 years highlighted the power of combining informant reports to predict child outcomes across a number of domains (Verhulst et al., 1994). When evaluating the predictive and incremental validity of these measures, it is important to consider the criterion variables used in each analysis, and the context in which they reference child behaviour, such as the home, education or care setting. When criterion measures are mapped across multiple settings, it is easier to achieve the ideal model of child assessment in which child wellbeing is captured from multiple perspectives (De Los Reyes et al., 2023; Makol et al., 2020). There is promising work emerging which examines the use of screening measures with preschool children through parent, educator and child psychologists' ratings, highlighting the value of gaining multifaceted assessments of child wellbeing, but there remains an opportunity to research this further (Gustafsson & Sund Levander, 2024).

These results contribute to the evidence that whilst educators often have minimal MH training, their knowledge of a referent classroom group of children may be advantageous in their assessment of child wellbeing and development, given strong clinical utility for educator-report. Future research should focus on developing and evaluating UMHS measures, with high screening accuracy and clinical utility for use in preschool populations by educators.

Examining the effectiveness of UMHS adds to the weight of evidence for multi-informant screening because in the context of UMHS, improved identification is clearly linked to increased diagnosis, referral and/or treatment and thus, improved outcomes for children. Yet our review found no studies reporting on the effectiveness of multi-informant raters. This follows a consistent pattern of omission in the literature in which effectiveness data, in terms of screening outcomes or the consequences of screening, are rarely reported adequately or at all (Houri & Miller, 2020). There is a distinct lack of data reporting screening effectiveness and limited research measuring the effectiveness of UMHS in the child MH space (Brinley et al., 2024). In the case of

screening amongst preschool children, none of the included studies investigated the effectiveness of multi-informant screening. This paucity of research remains despite multiple calls that we need to better understand the outcomes and consequences of screening. By effectively understanding outcomes, we can better understand the impact of UMHS (Houri & Miller, 2020). Thus, there is a clear need for future research to examine the effectiveness of screening in preschool populations with additional longitudinal studies measuring MH outcomes.

Given multi-informant report has long been the hallmark of developmental research and despite previous calls to action from highly cited research (e.g.,Achenbach et al., 1987; De Los Reyes et al., 2015), there remains a substantial gap in the study of multi-informant UMHS measures. There is a clear need for more research to demonstrate the incremental value of combining parent and educator reports in UMHS. Given demands on educators' and parents' time, lack of MH training for preschool educators and limited MH resources available to educators, the value of combined contributions needs to be made explicit. By demonstrating the predictive power of multi-informant reports, researchers can build the case for UMHS and improve child socio-emotional outcomes.

Critics have raised concerns that UMHS is not warranted in young children as it stigmatises them (Frances, 2012; Jureidini & Raven, 2012), and is neither wanted by parents nor educators (Humphrey & Wigelsworth, 2016). However, there is a mounting case refuting these conjectures. All three studies included in this review that investigated acceptability of UMHS reported strong support from parents and educators, for either specific screening measures or UMHS more generally in the preschool-age population. Importantly, each of these respondents evaluated the acceptability of UMHS after having completed screening for their children. A recent review of parent UMHS for children found that parents generally regarded UMHS as appropriate, across a range of contexts including schools, paediatric hospital settings, and well-child visits, although not all studies required parents to actually complete UMHS for their child (Brinley et al., 2024). The findings of the current review, therefore, add to existing research indicating that UMHS is, in fact, considered highly acceptable. Overall, the positive support from parents and educators for screening provides further evidence for the contention that preschools are ideal settings for screening, as they are accessible, less encumbered by MH stigma and a natural fit given parents frequently turn to educators for MH advice for their children (Desta et al., 2017; Pavuluri et al., 1996).

Whilst the studies in this review largely focussed on the acceptability of specific measures, there is a need to review screening more generally and in the current context. It may be possible that parents and educators have a growing awareness and understanding of early intervention for supporting child wellbeing, particularly in the aftermath of the global pandemic that had and continues to have significant ramifications on the wellbeing of children. More qualitative and quantitative research should be conducted in this area to identify perceived barriers, facilitators and perceived acceptability of multi-informant screening of preschool-aged children.

Strengths and Limitations

This review was the first to identify multi-informant UMHS measures utilised by parents and educators in preschool child populations and the first to evaluate their clinical utility, effectiveness and acceptability. However, there are a number of limitations that need to be kept in mind when interpreting the results. Studies that were excluded from this review included those that were not published in English, did not report predictive validity, effectiveness or acceptability data or were not peer reviewed. The exclusion criteria for child age may have limited the results and inadvertently excluded studies that included preschool-aged populations. It was often difficult to discern if 3–5 year old children were included in studies with broad-age range samples. The ages of "preschool" children were not always specified and differed between states or countries, as the age of school entry differed across jurisdictions. Secondary source or missing data were not sought from any authors to provide clarification. Moreover, some aspects of the reviewed literature were difficult to synthesise because of the disparate measures utilised. Many studies reported on author-developed measures, with varying identification methods, scoring and subscale outcomes, making synthesis difficult. Finally, risk of bias presents a possible limitation in the area of recruitment bias across the included studies. In evaluating the methodological quality of included studies using the MMAT, some studies may have been biased by not recruiting a representative sample of the target population (Hong et al., 2018). Given UMHS is intended for use amongst general, non-clinical populations in the community, it is of utmost importance that screening measures are tested in samples which are as representative as possible. The authors suggest that recruitment methodology should match implementation of UMHS in practice in order to minimise confounding variables in testing the efficacy of universal screening of whole populations. For example, UMHS measures should be implemented universally with as few participant eligibility criteria as reasonable and without interference from non-systematic factors such as teacher-nomination.

Implications

Parents and educators of preschool-aged children are integral to facilitating help-seeking for young children in need of MH support. They have a central role in early identification of MH problems and are a primary part of the pathway to enabling a child to access healthcare practitioners (De Los Reyes et al., 2015). However, based on the findings of this review, there are few measures that can be recommended to professionals in preschool and early childhood health settings for use with preschool-aged children. Based on the clinical utility of the included measures, the PBSS was the only measure to reach benchmark standards for predictive validity for both educator and parent report, however, the measure has not been published or made publicly available (Kettler et al., 2017; Owens et al., 2015). Having demonstrated incremental validity, the patented BASC-3, BESS and the freely available PSC-17 warrant further investigation to determine whether their multi-informant predictive validity are found to be acceptable. It is important to note that the widely used SDQ did not meet acceptable standards for predictive validity and as such, requires further research prior to being used for UMHS purposes in preschool children.

Based on the limited research examining incremental validity of preschool UMHS measures, it would appear that educator ratings are important and add significant incremental value to identifying children where problem behaviour occurs across multiple settings. Given how resource-intensive the implementation of multi-informant UMHS is, there is a strong need for further evidence that examines the incremental validity and effectiveness of UMHS. Future studies that investigate incremental validity need to carefully select gold-standard comparison measures which span various child contexts so that one informant is not inadvertently favoured over others (Haynes & Lench, 2003).

An important implication for future research is the examination of measures in terms of their ease of implementation and use in clinical practice. Some research in the area of measurement-based care has differentiated between the constructs of clinical utility and validity, defining "clinical utility" in terms of ease of implementation and usefulness for clinicians, and "validity" in terms of the evidence for specific measures (McLeod et al., 2022). If adopted in UMHS, this distinction may be of value as the ease of implementation and usefulness for clinicians and non-clinicians such as educators may be an important factor as to whether UMHS is more widely implemented.

Moreover, there is a need for additional studies examining UMHS, especially those which are representative and broader in scope. Large, population-level studies can provide important normative data that are inclusive of different cultural groups, children in urban and regional contexts and validated across ages. Thus, the results of this review suggest there continues to be a need to examine existing measures and develop new multi-informant UMHS measures for preschool-aged children that are psychometrically valid and reliable, accurate for identifying children at risk of MH difficulties. Such research should also enable further examination of acceptability of UMHS in preschool settings.

Conclusion

In examining the available UMHS measures utilised in preschool-aged children by parents and educators and by evaluating the psychometric evidence presented in studies, this review found promising results in terms of the clinical utility of measures to identify child MH concerns. However, with few measures reporting on all constructs of interest, there remains a need for further research on clinical utility, effectiveness and acceptability of measures. Accurate identification of young children at risk of MH problems is critical to providing early intervention and disrupting the trajectory of MH problems, and research to identify valid screening measures that can be easily implemented by teachers and educators is an important first step.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10567-024-00506-2.

Author contributions RM designed the study, developed the search strategy and performed the search, screened and extracted the data, conducted risk of bias assessment, analysed and interpreted the data and drafted the manuscript. LT was involved in the conception and the design of the review, screened and extracted the data, conducted risk of bias assessment, and reviewed the manuscript. SB screened and extracted the data, conducted risk of bias assessment, and drafted parts of the results including tables and figures, and supplementary files. TC screened and extracted the data, conducted risk of bias assessment. AT screened and extracted the data, conducted risk of bias assessment. JN screened and extracted the data, conducted risk of bias assessment. MD was involved in the conception and the design of the review, and reviewed the manuscript. All authors contributed to the development of the research questions and selection criteria; read, reviewed and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles* (Vol. 30). Burlington, VT: University of Vermont, Research center for children, youth

Achenbach, T. M. (1999). The Child Behavior Checklist and related instruments. *The use of psychological testing for treatment planning and outcomes assessment* (pp. 429–466). Lawrence Erlbaum Associates Publishers.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/Adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232. https://doi.org/10.1037/0033-2909.101.2.213

Alexander, K. E., & Mazza, D. (2015). Scrapping the healthy kids check: a lost opportunity. *The Medical Journal of Australia, 203*(8), 321–322.

Anderson, J. K., Ford, T., Soneson, E., Coon, J. T., Humphrey, A., Rogers, M., Moore, D., Jones, P. B., Clarke, E., & Howarth, E. (2019). A systematic review of effectiveness and cost-effectiveness of school-based identification of children and young people at risk of, or currently experiencing mental health difficulties. *Psychological Medicine, 49*(1), 9–19. https://doi.org/10.1017/S0033291718002490

Andrews, G. (1999). Efficacy, effectiveness and efficiency in mental health service delivery. *Australian & New Zealand Journal of Psychiatry, 33*(3), 316–322. https://doi.org/10.1046/j.1440-1614.1999.00581.x

Australian Bureau of Statistics. (2021). *Preschool Education, Australia: 2021*. https://www.abs.gov.au/statistics/people/education/preschool-education-australia/2021

Aylward, G. P. (1997). Conceptual issues in developmental screening and assessment. *Journal of Developmental & Behavioral Pediatrics, 18*(5), 340–349.

Barbarin, O. A. (2007). Mental health screening of preschool children: validity and reliability of ABLE. *The American Journal of Orthopsychiatry, 77*(3), 402–418.

Bayer, J., Hiscock, H., Scalzo, K., Mathers, M., McDonald, M., Morris, A., Birdseye, J., & Wake, M. (2009). Systematic review of preventive interventions for children's mental health: What would work in Australian contexts? *Australian and New Zealand Journal of Psychiatry, 43*(8), 695–710. https://doi.org/10.1080/00048670903001893

Becker-Haimes, E. M., Tabachnick, A. R., Last, B. S., Stewart, R. E., Hasan-Granier, A., & Beidas, R. S. (2020). Evidence base update for brief, free, and accessible youth mental health measures. *Journal of Clinical Child & Adolescent Psychology, 49*(1), 1–17. https://doi.org/10.1080/15374416.2019.1689824

Brinley, S., Tully, L. A., Carl, T., McLean, R. K., Cowan, C. S. M., Hawes, D. J., Dadds, M. R., & Northam, J. C. (2024). Universal child mental health screening for parents: A systematic review of the evidence. *Prevention Science*. https://doi.org/10.1007/s11121-024-01693-8

Casale, G., & Reyes, A. D. L. (2023). An integrated model for school-based mental health assessment in inclusive education. *Empirische Sonderpädagogik, 15*(3), 214–233.

Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. *Archives of General Psychiatry, 53*, 1033–1039. https://doi.org/10.1001/archpsyc.1996.01830110071009

Croft, S., Stride, C., Maughan, B., & Rowe, R. (2015). Validity of the strengths and difficulties questionnaire in preschool-aged children. *Pediatrics (Evanston), 135*(5), e1210–e1219. https://doi.org/10.1542/peds.2014-2920

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858–900. https://doi.org/10.1037/a0038498

De Los Reyes, A., Wang, M., Lerner, M. D., Makol, B. A., Fitzpatrick, O. M., & Weisz, J. R. (2023). The operations triad model and youth mental health assessments: Catalyzing a paradigm shift in measurement validation. *Journal of Clinical Child & Adolescent Psychology, 52*(1), 19–54. https://doi.org/10.1080/15374416.2022.2111684

DeLucia, E. A., Dike, J., Andrzejewski, T. M., & McDonnell, C. G. (2022). Autism screening practices in preschools and early childcare centers: A systematic review. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-022-00323-9

Desta, M., Deyessa, N., Fish, I., Maxwell, B., Zerihun, T., Levine, S., Fox, C., Giedd, J., Zelleke, T. G., Alem, A., & F. Garland, A. (2017). Empowering preschool teachers to identify mental health problems: A task-sharing intervention in Ethiopia. *Mind, Brain and Education, 11*(1), 32–42. https://doi.org/10.1111/mbe.12135

Doove, B., Feron, J., Feron, F., van Os, J., & Drukker, M. (2019). Validation of short instruments assessing parental and caregivers' perceptions on child health and development for personalized prevention. *Clinical Child Psychology and Psychiatry, 24*, 608–630. https://doi.org/10.1177/1359104518822673

Eklund, K., & Dowdy, E. (2014). Screening for behavioral and emotional risk versus traditional school identification methods. *School Mental Health: A Multidisciplinary Research and Practice Journal, 6*(1), 40–49. https://doi.org/10.1007/s12310-013-9109-1

Ellingson, K. D., Briggs-Gowan, M. J., Carter, A. S., & Horwitz, S. M. (2004). Parent identification of early emerging child behavior problems: Predictors of sharing parental concern with health providers. *Archives of Pediatrics & Adolescent Medicine, 158*(8), 766–772. https://doi.org/10.1001/archpedi.158.8.766

von der Embse, N., & De Los Reyes, A. (2024). Advancing equity in access to school mental health through multiple informant decision-making. *Journal of School Psychology, 104*, 101310. https://doi.org/10.1016/j.jsp.2024.101310

Ezpeleta, L., Granero, R., de la Osa, N., Penelo, E., & Domenech, J. M. (2013). Psychometric properties of the strengths and difficulties questionnaire(3–4) in 3-year-old preschoolers. *Comprehensive Psychiatry, 54*, 282–291. https://doi.org/10.1016/j.comppsych.2012.07.009

Feeney-Kettler, K. A., Kettler, R. J., & Dembitzer, L. (2019). Application of a multitrait-multimethod matrix to social, emotional, and behavioral screening. *Journal of Applied School Psychology, 35*(1), 20–36. https://doi.org/10.1080/15377903.2018.1468843

Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2011). Identification of preschool children at risk for emotional and behavioral disorders: Development and validation of a universal screening system. *Journal of School Psychology, 49*, 197–216. https://doi.org/10.1016/j.jsp.2011.02.001

Frances, A. (2012). Better safe than sorry. *Australian & New Zealand Journal of Psychiatry, 46*(8), 695–696. https://doi.org/10.1177/0004867412453775

Gall, G., Pagano, M. E., Desmond, M. S., Perrin, J. M., & Murphy, J. M. (2000). Utility of psychosocial screening at a school-based health center. *Journal of School Health, 70*(7), 292–298. https://doi.org/10.1111/j.1746-1561.2000.tb07254.x

Girio-Herrera, E., Dvorsky, M. R., & Owens, J. S. (2015). Mental health screening in kindergarten youth: a multistudy examination of the concurrent and diagnostic validity of the impairment rating scale. *Psychological Assessment, 27*, 215–227. https://doi.org/10.1037/a0037787

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135.

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(4), 400–403. https://doi.org/10.1097/CHI.0b013e3181985068

Gustafsson, B. M., & Sund Levander, M. (2024). The assessment of preschool children with ESSENCE symptoms: Concordance between parents, preschool teachers and child psychologists. *BMC Pediatrics, 24*(1), 191. https://doi.org/10.1186/s12887-024-04693-3

Harrison, J. R., Vannest, K. J., & Reynolds, C. R. (2013). Social acceptability of five screening instruments for social, emotional, and behavioral challenges. *Behavioral Disorders, 38*(3), 171–189.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15*(4), 456.

Heckman, J. J. (2008). The case for investing in disadvantaged young children. *Cesifo DICE Report, 6*(2), 3–8.

Hiscock, H., Mulraney, M., Efron, D., Freed, G., Coghill, D., Sciberras, E., Warren, H., & Sawyer, M. (2020). Use and predictors of health services among Australian children with mental health problems: A national prospective study. *Australian Journal of Psychology, 72*(1), 31–40. https://doi.org/10.1111/ajpy.12256

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT) version 2018* (Canada Patent No. 1148552). Canadian Intellectual Property Office - Industry Canada. https://content.iospress.com/articles/education-for-information/efi180221

Houri, A. K., & Miller, F. G. (2020). A systematic review of universal screeners used to evaluate social-emotional and behavioral aspects of kindergarten readiness. *Early Education and Development, 31*(5), 653–675.

Humphrey, N., & Wigelsworth, M. (2016). Making the case for universal school-based mental health screening. *Emotional and Behavioural Difficulties, 21*(1), 22–42. https://doi.org/10.1080/13632752.2015.1120051

Veritas Health Innovation. *Covidence systematic review software*. In www.covidence.org

Jellinek, M. S., Murphy, J. M., Robinson, J., Feins, A., Lamb, S., & Fenton, T. (1988). Pediatric symptom checklist: Screening school-age children for psychosocial dysfunction. *The Journal of Pediatrics, 112*(2), 201–209.

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health, 1971*(105), 2283–2290. https://doi.org/10.2105/AJPH.2015.302630

Jureidini, J., & Raven, M. (2012). Healthy kids check: Lack of transparency and misplaced faith in the benefits of screening. *Australian and New Zealand Journal of Psychiatry, 46*(10), 924–927. https://doi.org/10.1177/0004867412460731

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 67*(2), 342–356. https://doi.org/10.1177/0013164407067002l001

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry, 159*(9), 1548–1555.

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the strengths and difficulties questionnaire. *International Journal of Behavioral Development, 40*(1), 64–75. https://doi.org/10.1177/0165025415570647

Kettler, R. J., Feeney-Kettler, K. A., & Dembitzer, L. (2017). Social, emotional, and behavioral screening: A comparison of two measures and two methods across informants. *Journal of School Psychology, 64*, 93–108. https://doi.org/10.1016/j.jsp.2017.05.002

Konold, T. R., Walthall, J. C., & Pianta, R. C. (2004). The behavior of child behavior ratings: Measurement structure of the child behavior checklist across time, informants, and child gender. *Behavioral Disorders, 29*(4), 372–383. https://doi.org/10.1177/019874290402900405

Kowalenko, N. M. (2012). The three year old social and emotional well-being check is good news. *Australian and New Zealand Journal of Psychiatry, 46*(10), 928–931. https://doi.org/10.1177/0004867412461534

Lavigne, J. V., Lebailly, S. A., Hopkins, J., Gouze, K. R., & Binns, H. J. (2009). The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *Journal of Clinical Child and Adolescent Psychology, 38*(3), 315–328. https://doi.org/10.1080/15374410902851382

Lavigne, J. V., Meyers, K. M., & Feldman, M. (2016). Systematic review: Classification accuracy of behavioral screening measures for use in integrated primary care settings. *Journal of Pediatric Psychology, 41*(10), 1091–1109. https://doi.org/10.1093/jpepsy/jsw049

Makol, B. A., Youngstrom, E. A., Racz, S. J., Qasmieh, N., Glenn, L. E., & De Los Reyes, A. (2020). Integrating multiple informants' reports: How conceptual and measurement models may address long-standing problems in clinical decision-making. *Clinical Psychological Science, 8*(6), 953–970. https://doi.org/10.1177/2167702620924439

McLeod, B. D., Jensen-Doss, A., Lyon, A. R., Douglas, S., & Beidas, R. S. (2022). To utility and beyond! Specifying and advancing the utility of measurement-based care for youth. *Journal of Clinical Child & Adolescent Psychology, 51*(4), 375–388.

Moore, S. A., Dowdy, E., Fleury, I., DiStefano, C., & Greer, F. W. (2022a). Comparing informants for mental health screening at the preschool level. *School Psychology Review, 51*, 589–608. https://doi.org/10.1080/2372966x.2020.1841546

Moore, S. A., Dowdy, E., Hinton, T., DiStefano, C., & Greer, F. W. (2022b). Moving towards implementation of universal mental health screening by examining attitudes toward school-based practices. *Behavioral Disorders, 47*, 166–175. https://doi.org/10.1177/0198742920982591

Murphy, J. M., Ichinose, C., Hicks, R. C., Kingdon, D., Crist-Whitzel, J., Jordan, P., Feldman, G., & Jellinek, M. S. (1996). Utility of the pediatric symptom checklist as a psychosocial screen to meet the federal early and periodic screening, diagnosis, and treatment (EPSDT) standards: A pilot study. *Journal of Pediatrics, 129*(6), 864–869. https://doi.org/10.1016/s0022-3476(96)70030-6

National Center for Education Statistics. (2016, 14/6/2023). *National Household Education Surveys Program*. National Center for Education Statistics. https://nces.ed.gov/nhes/tables/ECPP_HoursPerWeek_Care.asp

National Center for Education Statistics. (2018, 14/6/2023). *Digest of Education statistics, 2018*. National Center for Education Statistics. https://nces.ed.gov/programs/digest/d18/tables/dt18_202.10.asp

Oh, E., Mathers, M., Hiscock, H., Wake, M., & Bayer, J. (2015). Professional help seeking for young children with mental health problems. *Australian Journal of Psychology, 67*(3), 187–195. https://doi.org/10.1111/ajpy.12072

Owens, J. S., Storer, J., Holdaway, A. S., Serrano, V. J., Watabe, Y., Himawan, L. K., Krelko, R. E., Vause, K. J., Girio-Herrera, E., & Andrews, N. (2015). Screening for social, emotional, and behavioral problems at kindergarten entry: Utility and incremental validity of parent report. *School Psychology Review, 44*, 21–40. https://doi.org/10.17105/SPR44-1.21-40

Pavuluri, M. N., Luk, S.-L., & McGee, R. O. B. (1996). Help-seeking for behavior problems by parents of preschool children: A community study. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*, 215–222. https://doi.org/10.1097/00004583-199602000-00015

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53*(2), 109.

Reardon, T., Harvey, K., Baranowska, M., O'Brien, D., Smith, L., & Creswell, C. (2017). What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? A systematic review of qualitative and quantitative studies. *European Child & Adolescent Psychiatry, 26*(6), 623–647. https://doi.org/10.1007/s00787-016-0930-6

Rhodes, A. (2017). *Child mental health problems: Can parents spot the signs?* T. R. C. s. Hospital. https://www.rchpoll.org.au/polls/child-mental-health-problemscan-parents-spot-the-signs/

Robinson, M., Oddy, W. H., Li, J., Kendall, G. E., De Klerk, N. H., Silburn, S. R., Zubrick, S. R., Newnham, J. P., Stanley, F. J., & Mattes, E. (2008). Pre-and postnatal influences on preschool mental health: A large-scale cohort study. *Journal of Child Psychology and Psychiatry, 49*(10), 1118–1128. https://doi.org/10.1111/j.1469-7610.2008.01955.x

Sawyer, A. C. P., Chittleborough, C. R., Lynch, J. W., Baghurst, P., Mittinty, M. N., Kaim, A. L. E., & Sawyer, M. G. (2014). Can screening 4–5 year olds accurately identify children who will have teacher-reported mental health problems when children are aged 6–7 years? *Australian & New Zealand Journal of Psychiatry, 48*(6), 554–563. https://doi.org/10.1177/0004867413514491

Schnyder, N., Sawyer, M. G., Lawrence, D., Panczak, R., Burgess, P., & Harris, M. G. (2020). Barriers to mental health care for Australian children and adolescents in 1998 and 2013–2014. *Australian & New Zealand Journal of Psychiatry, 54*(10), 1007–1019. https://doi.org/10.1177/0004867420919158

Schubiner, H., Tzelepis, A., Wright, K., & Podany, E. (1994). The clinical utility of the safe times questionnaire. *Journal of Adolescent Health, 15*, 374–382. https://doi.org/10.1016/1054-139X(94)90260-7

Scott, S., Briskman, J., & Dadds, M. R. (2011). *Measuring parenting in community and public health research using brief child and parent reports* https://doi.org/10.1007/s10826-010-9398-z]. Springer. https://link.springer.com/content/pdf/https://doi.org/10.1007/s10826-010-9398-z.pdf

Sim, F., Thompson, L., Marryat, L., Ramparsad, N., & Wilson, P. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review [Report]. *PLoS ONE, 14*, e0211409.

Smith, S. R. (2007). Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *Journal of Psychoeducational Assessment, 25*(2), 139–149. https://doi.org/10.1177/0734282906296233

Stefan, C. A., & Miclea, M. (2017). Reliability and validity of two brief screening measures of preschoolers' social-emotional competencies. *School Mental Health, 9*(1), 44–65. https://doi.org/10.1007/s12310-016-9200-5

Stormont, M., & Stebbins, M. S. (2005). Preschool teachers' knowledge, opinions, and educational experiences with attention deficit/hyperactivity disorder. *Teacher Education and Special Education, 28*(1), 52–61. https://doi.org/10.1177/088840640502800106

Takayanagi, N., Yoshida, S., Yasuda, S., Adachi, M., Kaneda-Osato, A., Tanaka, M., Masuda, T., Kuribayashi, M., Saito, M., & Nakamura, K. (2016). Psychometric properties of the Japanese ADHD-RS in preschool children. *Research in Developmental Disabilities, 55*, 268–278. https://doi.org/10.1016/j.ridd.2016.05.002

Verhulst, F. C., Koot, H. M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. *Journal of Abnormal Child Psychology, 22*(5), 531–546. https://doi.org/10.1007/bf02168936

Wakschlag, L. S., Briggs-Gowan, M. J., Carter, A. S., Hill, C., Danis, B., Keenan, K., McCarthy, K. J., & Leventhal, B. L. (2007). A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *Journal of Child Psychology and Psychiatry, 48*(10), 976–987.

ANZJP

Australian & New Zealand Journal of Psychiatry
1–11
DOI: 10.1177/00048674251342952

Reliability, predictive validity and normative data for the Pediatric Symptom Checklist-17 in a national Australian sample

Rebecca K McLean, Lucy A Tully, Mark R Dadds; Growing Minds Australia-Clinical Trials Network

Abstract

Background: We evaluated the reliability, validity and acceptability of the Pediatric Symptom Checklist-17, a free, brief measure of child mental health, in a sample of parents of preschool-age (3–5 years) and school-age children (6–17 years). This is the first study to examine parent-reported Pediatric Symptom Checklist-17 for children aged 3 years.

Method: A national community sample of Australian parents ($N = 2097$) completed a demographic questionnaire and the Pediatric Symptom Checklist-17. We used a cross-sectional, test–retest design to assess the structural validity, internal consistency, test–retest reliability, concurrent and predictive validity of the Pediatric Symptom Checklist-17 total and subscale scores. Predictive validity was evaluated using a sub-sample of parents ($n = 122$) who completed the Pediatric Symptom Checklist-17 and Child Behaviour Checklist at a second timepoint. Normative data were also produced.

Results: Factor analysis supported the three-factor model for the Pediatric Symptom Checklist-17. Total ($r = 0.82$–0.93) and subscale scores ($r = 0.75$–0.89) strongly correlated with the Child Behaviour Checklist and demonstrated strong internal consistency (total scores $\alpha = 0.87$–0.88). Test–retest reliability was acceptable for school-age children. The Pediatric Symptom Checklist-17 demonstrated excellent classification accuracy for preschool children; however, it did not perform as strongly in older age ranges. Normative data for age and gender were produced for the measure. Results indicated high levels of parent acceptability for the measure.

Conclusion: Findings contribute new validation evidence for the use of the Pediatric Symptom Checklist-17 as a screening and assessment measure in research and clinical settings, for parents of children from 3 years and above. This study is the first to ascertain Australian normative data for the Pediatric Symptom Checklist-17.

Keywords

Child mental health, universal screening, assessment, psychometrics, clinical psychology

While it has been established that many mental health (MH) problems begin before 14 years of age (Kessler et al., 2007), a recent review of prospective longitudinal studies found MH difficulties experienced in children as young as 5 years are associated with adult MH disorders (Mulraney et al., 2021). This highlights the need to identify preschool children as a cohort of focus due to parents' lack of confidence in recognising MH difficulties, high proportion of children with clinically significant MH problems and the high levels of unmet MH needs in this age range (Kataoka et al., 2002; Oh et al., 2015; Rhodes, 2017; Robinson et al., 2008).

A plethora of child MH measures currently exist, which may assist parents to identify child MH difficulties. However, there is a lack of available, brief, parent-report measures for screening and assessment that are appropriate for a range of ages including preschool children aged

School of Psychology, The University of Sydney, Sydney, NSW, Australia

Corresponding author:
Rebecca K McLean, School of Psychology, The University of Sydney, 97 Church Street, Camperdown NSW 2050, Australia.
Email: rebecca.mclean@sydney.edu.au

3–5 years (McLean et al., 2024; Tully et al., 2024). Measures also need to produce valid and reliable scores (Humphrey and Wigelsworth, 2016). In addition, clinical cut-offs and normative data relevant to specific populations can facilitate the use of measures in research and clinical practice (Kendall and Sheldrick, 2000). The prevalence of child MH problems in Australia is high and stable (Sawyer et al., 2018), despite the increased availability of evidence-based intervention. There is an urgent need to identify measures that can be used to detect children with MH problems in the Australian context. Thus, the aim of this study was to validate and ascertain normative data for the Pediatric Symptom Checklist-17 (PSC-17) in a national Australian sample.

Overview of the PSC-17

Adapted from a 35-item version (Jellinek and Murphy, 1988), the PSC-17 is a 17-item, parent-report measure of child MH, validated and normed for use with children aged 4–17 years in the United States (Gardner et al., 1999). Total scores and three subscales based on internalising, externalising and attention problems are produced by summing responses to items. Items are based on a reflective model and response options include *Never, Sometimes, Often*. It has been one of the most frequently recommended screening measures by US MH authorities (Semansky et al., 2003). Within Australia, interest in the PSC-17 has been growing due to the increasing use of digitised measures in clinical and research settings and lack of freely available measures for assessing parent reports of child MH (Tully et al., 2024).

Widely used, the PSC-17 is considered feasible to implement in primary care settings due to its advantages over other measures; it is brief, free to implement in digital and paper-based format, easy to comprehend, simple to score, covers a wide range of ages and is also suitable for use as a screening measure by caregivers who may have limited or no training in MH, such as early childhood educators and teachers (Sheldrick et al., 2012).

Psychometric properties

The technical accuracy of measures, as indicated by various psychometric or measurement properties, is critical when considering the use of measures in screening and assessment of child MH. The PSC-17 has established validity and reliability in children aged 4–17 years, and as part of a recent review of 95 child MH measures, was rated as having excellent psychometric properties in terms of reliability, validity, normative data and treatment sensitivity; one of 21 measures to be rated as such (Becker-Haimes et al., 2020; Spencer et al., 2020; Stoppelbein et al., 2012). Multiple studies have reported strong internal consistency for the PSC-17; Cronbach's alpha 0.85–0.89 for total scores, 0.78–0.79 internalising, 0.80–0.83 externalising and 0.82–0.83 attention (Gardner et al., 1999; Murphy et al., 2016; Wagner et al., 2015).

There is limited research on the test–retest reliability of the PSC-17. One community study reported good test–retest reliability (intraclass coefficient [ICC] = 0.85) after 8–14 days (Murphy et al., 2016). Another study reported an ICC of 0.55; however, the retest was implemented at 6 months, among children in foster care (Jacobson et al., 2019).

Previous research has reported the PSC-17 subscales had significant relationships with associated diagnosed disorders, e.g., internalising subscale scores with depression and anxiety, indicating construct and convergent validity (Jacobson et al., 2019). The original validation study for the PSC-17 found total scores correlated with the clinician-rated Children's Global Assessment Scale ($r = -0.64$; 95% confidence interval [CI]: −0.71 to −0.55) and Child Behaviour Checklist (CBCL) ($r = -0.60$; 95% CI: −0.67 to −0.52) (Gardner et al., 2007).

The PSC-17 has strong predictive validity, which is important for screening measures. A systematic review of screening measures identified three studies reporting high sensitivity (SE) 0.82–0.95 and specificity (SP) 0.81–0.91 (Lavigne et al., 2016). Mean sensitivity was 0.90 and specificity 0.85, which surpassed benchmark standards for developmental screening measures (Aylward, 1997; Lavigne et al., 2016).

While the PSC-17 has been well-validated in older children, there remains a need to evaluate the PSC-17 as a measure of child MH in young children in terms of factor structure, internal consistency, test–retest reliability, concurrent and predictive validity.

Use of the PSC-17 with preschool-aged children

While much of the existing research on the PSC-17 has been on school-aged children, limited research has examined the predictive validity and reliability of the PSC-17 in preschools. Studies with teacher report found strong internal consistency across total scores and subscales (DiStefano et al., 2017; Liu et al., 2020; Moore et al., 2021). In a large paired-sample study, partial scalar invariance was established indicating parents and teachers of children aged 4–5 years interpreted items similarly (Gao et al., 2022). However, there remains a need to validate parent-reported ratings in children aged 3 years, since the measure is currently being used in clinical practice with this age (Wagner et al., 2015; Benheim T, Jellinek M and Murphy M, Personal communication, March 7, 2023).

Normative data

The PSC-17 has not been validated with parents in Australia nor is there normative data outside of the United States,

which is important for the systematic interpretation of results and efficient identification of children who may have emerging MH problems (Sheldrick et al., 2012). Moreover, the PSC-17 does not have normative data for the preschool age in any country. Using normative data developed for older children, in clinical settings, may not be appropriate for young children, in community settings. US data also may not be appropriate in the Australian cultural context. Producing localised normative data ensures recommended clinical cut-offs are appropriate for screening and assessment purposes in non-clinical populations.

Acceptability

Some researchers and practitioners have considered the use of MH screening measures with young children as contentious (Frances, 2012; Jureidini and Raven, 2012). The (lack of) acceptability of a measure can affect uptake and implementation of child MH or screening measures across settings (Harrison et al., 2013). Acceptability may be defined by its perception as helpful or useful, whether it is easy to understand and whether users would recommend it to others (Brinley et al., 2024; Humphrey and Wigelsworth, 2016). Therefore, the perceived acceptability of a measure by users (in this case parents) is an important aspect to investigate if it is to be adopted as part of improving access to early intervention.

The current study

This study seeks to add to existing evidence by evaluating the parent-reported PSC-17 and ascertaining normative data across preschool and school-age populations. To our knowledge, this is the first study to validate the PSC-17 for children aged 3 years and the first to ascertain normative data, including age and gender versions for an Australian population. The aims of the study were, first, to evaluate the psychometric properties of the PSC-17 and validate parent ratings for preschool children (3–5 years) and school-age children (6–17 years), in terms of factor structure, internal consistency, test–retest reliability and concurrent and predictive validity. Second, to ascertain the specific age and gender normative data for parent-report PSC-17 in Australia for children aged 3–17 years and finally to examine the acceptability among parents of preschool children compared to school children.

Method

Procedures

Study approval was obtained from the University of Sydney Human Research Ethics Committee. There are no set conventions specific to sample sizes for establishing normative data, albeit ideal samples should be large and representative of the defined population. Previous normative studies have recruited samples varying from 500 to 5400 participants (D'Souza et al., 2017; Mellor, 2005). Based on methods previously used by the investigators, this study recruited two samples of 1000 parents of preschool (3–5 years) and school-age children (6–17 years).

Participants were recruited through an online research panel and sampled to represent Australian families based on key sociodemographic characteristics from national census data (i.e. household income, marital status and residential location). Inclusion criteria included adult parents or caregivers of a child aged 3–17 years, who reside in Australia and have basic English literacy. Caregivers could include anyone in a caregiving role such as fathers, mothers, kinship carers or foster carers. Quotas were used for the recruitment of two samples, including an even ratio of child gender and ages within each sample. See supplementary material for further details about recruitment procedures.

After providing informed consent, participants completed measures which were counterbalanced in order of completion. Baseline questionnaires took approximately 10–15 minutes to complete. After data quality checks, the final sample included parent reports of 1048 preschool and 1049 school-age children.

To evaluate retest reliability, concurrent and predictive validity, a sub-sample of participants completed a second set of questionnaires, which took approximately 15–20 minutes to complete. A subset of preschool parents ($n=55$) completed questionnaires 6–73 days after baseline ($M=35$) and 67 school parents 11–69 days ($M=43$) later. In the preschool sample, 16.4% of parents and in the school sample, 4.5% of participants completed the second assessment between 7 and 14 days.

Participants

Preschool children were aged 3.87 years on average (SD = 0.78) and school children were aged 11.28 years on average (SD = 3.42). Preschool and school-age samples were almost evenly split between male and female (see Table 1). One preschool child and four school-aged children were identified as non-binary gender. Child ethnicity was predominantly Caucasian, followed by Aboriginal and Torres Strait Islander and multiple ethnicities (see Table 2). Parents from all states and territories were represented in both samples and were culturally diverse. See supplementary material for further details.

Measures

In addition to the PSC-17, parents completed questions about child and family sociodemographic details and the following measures.

CBCL. The CBCL (Achenbach, 1999) is a comprehensive measure of child MH with strong psychometric properties

across two age versions – 1.5–5 and 6–18 years. It is the most common criterion measure in classification accuracy studies (Lavigne et al., 2016). The CBCL consists of 99 items (1.5–5 years) and 113 items (6–18 years) and produces internalising, externalising and total problems scales. The CBCL was used in the current study as a criterion measure to establish concurrent and predictive validity.

Table 1. Age and Gender Distribution of the Sample.

Child age in years	Male *N*	Female *N*	Total *N*
3	209	186	395
4	205	187	393[a]
5	141	119	260
6	42	46	88
7	55	48	103
8	44	50	94
9	48	43	91
10	46	43	89
11	40	47	87
12	48	38	86
13	44	49	93
14	44	40	85[a]
15	29	45	76[a]
16	39	39	79[a]
17	35	43	78
Total	1069	1023	2097

[a]Five children were identified as non-binary gender.

Acceptability. Parents' views on the acceptability of the PSC-17 were assessed using four items adapted from a parent-report measure developed by Hawes et al. (2021). Responses were collected on a 5-point Likert-type response scale from *not at all/strongly disagree* (1) to *very/strongly agree* (5). See supplementary material for further details about this measure.

Analytic plan

Factor structure, reliability and validity were evaluated separately for preschool and school-age children. Using the original three-factor model (Gardner et al., 1999), confirmatory factor analysis (CFA) was performed and fit statistics reported. CFA was conducted using SPSS Amos (IBM SPSS, 2022). The latent variables, internalising, externalising and attention, were allowed to be correlated, and all measurement error was presumed to be uncorrelated. Hu and Bentler's (1999) standards of acceptable model fit were used to evaluate the model. Further details are outlined in the supplementary material.

To align with the age of school entry (preschool, primary and secondary school) and CBCL scoring, analyses for predictive validity were conducted on young children (3–5 years), older children (6–11 years) and adolescents (12–17 years). Receiver operator characteristic (ROC) analysis in each of these age groups examined the association between PSC-17 subscales and CBCL case classifications (clinical/non-clinical cases). Normative data are also

Table 2. Child Ethnicity.

	Children aged 3–5 years		Children aged 6–17 years	
Child ethnicity	*N*	Percent	*N*	Percent
Aboriginal/Torres Strait Islander	76	7.2	62	5.9
African/African American	5	0.5	8	0.8
Caucasian (e.g. British, European)	685	65.3	762	72.6
East Asian (e.g. Chinese, Japanese, Korean)	42	4.0	29	2.8
Hispanic/Latino	3	0.3	4	0.4
Middle-Eastern (e.g. Egyptian, Iraqi, Lebanese)	10	1.0	11	1
Polynesian (e.g. Pacific Islander, Māori)	12	1.1	6	0.6
South Asian (e.g. Indian, Pakistani)	53	5.0	37	3.5
South-East Asian (e.g. Vietnamese, Filipino)	45	4.3	24	2.3
Other or multiple ethnicities identified	117	11.2	106	10.1
Total	1048	100	1049	100

presented for these age ranges. As the online questionnaire required input for all items, there were no missing data.

Since the test–retest time interval was longer than intended, reliability estimates for short intervals (7–14 days) compared those outside that time period (<7 and >14 days) were analysed, and no significant difference was found. Thus, we report retest reliability for the full sub-sample.

Psychometric properties were evaluated against Youngstrom et al.'s (2017) criteria set out in a rubric by De Los Reyes and Langer (2018). Since the rubric does not include specific thresholds for predictive validity or classification accuracy, predictive validity of the scale was evaluated against benchmark criteria for developmental screening measures of 70% and above for sensitivity and specificity (Aylward, 1997), and area under curve (AUC) values of 0.80 or higher were considered 'very good' and 0.90 or higher as 'excellent' (Chaffin et al., 2017).

Normative data were determined by reporting scoring cut-offs for the 90th (borderline) and 95th percentile (at-risk cases) for subscales and total scores, gender and age. At-risk scores indicate positive identification on the screening measure and further investigation of child MH may be warranted.

Results

Factor structure. For the preschool sample, the chi-square fit statistic was large and significant, $\chi^2(116, N=1048)=661.73$, $p<0.001$. $\chi^2/df=5.70$. Other fit statistics, however, revealed a root mean square error of approximation (RMSEA)=0.07, CI 90%: 0.06–0.07, standardised root mean square residual (SRMR)=0.05, comparative fit index (CFI)=0.91, Tucker–Lewis index (TLI)=0.90, indicating some support for the hypothesised model. Factor loading estimates were moderately to strongly correlated to the latent factors (R^2: 0.49–0.82) in the preschool sample.

For the school sample, the chi-square fit statistic was large and significant, $\chi^2(116, N=1049)=474.84$, $p<0.001$. $\chi^2/df=4.09$. Other fit indices supported the hypothesised model in the school-age sample, RMSEA=0.05, CI 90%: 0.05–0.06, SRMR=0.05, CFI=0.94, TLI=0.93. Factor loading estimates were moderately to strongly correlated to the latent factors (R^2: 0.57–0.84) in the school sample.

The modified preschool model and the school model met three out of the six criteria, with two criteria 'close to' cut-offs.

Distribution of scores. The mean, standard deviation and ranges for each of the samples are outlined in Table 3.

Reliability. In the preschool sample, good internal consistency was found for the PSC-17 total scores ($\alpha=0.87$), externalising ($\alpha=0.83$) and attention subscales ($\alpha=0.84$). Internal consistency of the internalising subscale was adequate ($\alpha=0.73$). In the school sample, internal consistency for total scores ($\alpha=0.88$), externalising ($\alpha=0.83$) and attention subscales ($\alpha=0.84$) was also good. Internal consistency of the internalising subscale was adequate ($\alpha=0.78$).

In the preschool sample, analyses for test–retest reliability was performed and reported as follows: total scores (ICC=0.66), internalising (ICC=0.62), externalising (ICC=0.61) and attention subscales (ICC=0.63). In the school sample, good test–retest reliability was found for total scores (ICC=0.73) and the internalising subscale (ICC=0.70). Test–retest reliability for externalising (ICC=0.67) and attention subscales (ICC=0.68) did not meet adequate criteria. ICC estimates and their 95% confident intervals were calculated based on a single rating, absolute-agreement, two-way mixed-effects model.

Concurrent validity. The PSC-17 strongly correlated with the CBCL across ages demonstrating concurrent validity. In the preschool sample, PSC-17 total scores strongly and significantly correlated with the CBCL total problems at the bivariate level, $r=0.93$, $p<0.001$, $n=55$. PSC-17 subscales also strongly correlated with the CBCL subscales: internalising $r=0.86$, $p<0.001$; externalising $r=0.86$, $p<0.001$; and attention $r=0.89$, $p<0.001$. In the school sample, PSC-17 total scores were strongly associated with the CBCL total problems, $r=0.82$, $p<0.001$, $n=67$; internalising $r=0.75$, $p<0.001$; externalising $r=0.80$, $p<0.001$; and attention $r=0.81$, $p<0.001$.

In the preschool sample, the AUC for PSC-17 total scores was excellent 0.86 with a standard error of 0.12. This was significant at 0.002 (CI: 0.628–1.084). Internalising AUC was adequate 0.68, SE: 0.14, $p=0.184$ (CI: 0.414–0.947); externalising AUC was excellent 0.83, SE: 0.14, $p=0.018$ (CI: 0.555–1.101); and attention AUC was excellent 0.89, SE: 0.05, $p<0.001$ (CI: 0.792–0.996). In the school sample, the AUC for PSC-17 total scores was excellent 0.83 with a standard error of 0.07. This was significant at $p<0.001$ (CI: 0.683–0.967). Internalising AUC 0.79, SE: 0.07, $p<0.001$ (CI: 0.649–0.926); externalising AUC: 0.88, SE: 0.08, $p<0.001$ (CI: 0.713–1.038); and attention AUC: 0.91, SE: 0.08, $p<0.001$ (CI: 0.758–1.065) subscales all met criteria for excellent validity.

Predictive validity. Predictive validity was examined in young children (3–5 years, $n=55$), older children (6–11 years, $n=30$) and adolescents (12–17 years, $n=37$). A low number of clinical cases across ages meant assumptions were violated for chi-square tests. The older children and adolescent categories were combined for analysis; however, case numbers for the attention subscale remained low. In the preschool sample, high sensitivity was demonstrated for total scores (83.3), externalising (83.3) and attention subscales (77.8). Very high specificity was

Table 3. Mean scores and banding for PSC-17 subscales and total scores.

		3–5 years ($n = 1048$)	6–17 years ($n = 1049$)
Internalising	M	1.95	3.50
	SD	1.80	2.19
	Range	0–10	0–10
	Possible range	0–10	0–10
	α	0.73	0.78
	Top 5% of cases	⩾5	⩾7
	Top 10% of cases	⩾4	⩾6
Externalising	M	3.66	3.41
	SD	2.86	2.96
	Range	0–14	0–14
	Possible range	0–14	0–14
	α	0.83	0.83
	Top 5% of cases	⩾9	⩾9
	Top 10% of cases	⩾7	⩾7
Attention	M	3.12	3.68
	SD	2.59	2.78
	Range	0–10	0–10
	Possible range	0–10	0–10
	α	0.84	0.84
	Top 5% of cases	⩾8	⩾9
	Top 10% of cases	⩾7	⩾8
Total scores	M	8.73	10.59
	SD	6.04	6.30
	Range	0–34	0–32
	Possible range	0–34	0–34
	α	0.87	0.88
	Top 5% of cases	⩾20	⩾22
	Top 10% of cases	⩾17	⩾19

demonstrated for total scores and all subscales (93.3–97.8). In the school sample, moderate sensitivity was demonstrated for the externalising subscale only (71.4). High specificity was demonstrated for total scores and all subscales (88.7–100). Predictive validity of the PSC-17 using original clinical cut off scores for each of the samples is presented in Table 4. ROC analysis identified alternate cut off points which maximised sensitivity and specificity. Alternate clinical cut-points and their respective sensitivity and specificity are presented in Table 5.

Normative data. In the preschool-age range, 17% of cases were positively identified (at-risk) on total scores, 9.4% internalising, 16.5% externalising and 13.5% attention. In the school-age range (6–17 years), 25.3% were positively identified on total scores, 31.8% internalising, 15.1% externalising and 19.2% attention. Fine-grained age analysis revealed 25.4% of older children were positively identified on total scores, 26.1% internalising, 16.8% externalising and 20.5% attention. Among adolescents, 25.2% were positively identified on total scores, 38.2% internalising, 13.1% externalising and 17.7% attention. For each age and gender group, Table 6 presents cut-off scores for the highest-ranking cases.

Males across ages scored higher than females on total scores, as such, further analyses were performed to test the significances of these differences by gender. Preschool males ($M = 9.49$) scored significantly higher than females ($M = 7.89$) on total scores as indicated by an independent samples t-test ($F = 7.79$, $p < 0.001$). A two-way analysis of variance (ANOVA) indicated there was no significant

Table 4. Predictive Validity of PSC-17 using original clinical cut-off scores.

PSC-17	Cut-off score	AUC	Sensitivity %	Specificity %	PPV	NPV
3–5 years ($n=54$)						
Internalising	5	0.80 (CI: 0.632–0.977)	25	97.8	66.7	73.7
Externalising	7	0.82 (CI: 0.551–1.095)	83.3	95.8	71.4	97.9
Attention	7	0.87 (CI: 0.740–0.990)	77.8	93.3	70	95.5
Total score	15	0.86 (CI: 0.628–1.084)	83.3	93.8	62.5	97.8
6–17 years ($n=67$)						
Internalising	5	0.79 (CI: 0.649–0.926)	0	100	0	77.3
Externalising	7	0.88 (CI: 0.713–1.038)	71.4	91.5	50.0	96.4
Attention	7	0.91 (CI: 0.758–1.065)	0	100	0	92.4
Total score	15	0.83 (CI: 0.683–0.967)	61.5	88.7	57.1	75.8

Table 5. Predictive Validity of PSC-17 using alternate clinical cut off scores.

PSC-17	Cut-off score	Sensitivity %	Specificity %
3–5 years ($n=54$)			
Internalising	2.5	75	76.1
Externalising	6.5	83.3	95.8
Attention	3.5 or 7.5	100 or 66.7	58.3 or 91.7
Total score	17	83.3	95.8
6–11 years ($n=30$)			
Internalising	5	40	80
Externalising	0.5	50	92.3
Attention	0.5	100	92.6
Total score	16.5	60	96
12–17 years ($n=37$)			
Internalising	4.5	80	84.6
Externalising	6	100	90.9
Attention	2.5	100	55.9
Total score	11.5	87.5	85.7

interaction between the effects of preschool age and gender on PSC-17 total scores. While total scores were also higher for school males ($M=11.04$) than for females ($M=10.10$), this was not a significant difference ($F=2.98$, $p=0.085$).

Acceptability. Parents' ratings of dimensions related to the acceptability indicated that most parents found the PSC-17 to be very or quite appropriate (preschool 87.3%, $n=55$; school-age 82.9%, $n=1049$) and very or quite clear (preschool 92.8%; school-age 93.3%). Most parents were very or quite confident in the ratings they provided (preschool 96.3%; school-age 95.3%); and agreed or strongly agreed that the PSC-17 asked important questions about their child's MH (preschool 89.1%; school-age 92.5%). Independent *t*-tests showed that there was no significant difference between ratings of acceptability between preschool parents and school-age parents.

Discussion

The purpose of this study was to evaluate the reliability and validity of the parent-reported PSC-17 in young children, ascertain Australian normative data and examine acceptability among parents. To address the first aim, this study presents new evidence in support of the measure's

Table 6. Mean scores and banding for PSC-17 by age and gender.

		3–5 years		6–11 years		12–17 years	
		Male (*n* = 555)	Female (*n* = 492)	Male (*n* = 275)	Female (*n* = 277)	Male (*n* = 239)	Female (*n* = 254)
Internalising	M	2.04	1.85	3.46	3.01	3.64	3.86
	SD	1.87	1.71	2.09	1.76	2.47	2.31
	Range	0–10	0–9	0–10	0–9	0–10	0–10
	α	0.75	0.71	0.76	0.67	0.83	0.80
	PSC case rate (% positive)	10.1	8.7	31.3	20.9	36.8	39.0
	At-risk (top 5%)	6	5	7	6	8	8.25
	Borderline (top 10%)	5	4	6	5	7	7
Externalising	M	3.98	3.30	3.96	3.41	3.20	2.96
	SD	3.03	2.62	3.11	2.90	2.83	2.87
	Range	0–14	0–14	0–13	0–14	0–14	0–14
	α	0.84	0.80	0.84	0.83	0.80	0.83
	PSC case rate (% positive)	20.5	12.0	21.8	11.9	12.6	13.0
	At-risk (top 5%)	10	8	10	9	8	8
	Borderline (top 10%)	8	7	8	7	7	7
Attention	M	3.48	2.73	4.20	3.58	3.53	3.38
	SD	2.70	2.41	2.75	2.81	2.73	2.75
	Range	0–10	0–10	0–10	0–10	0–10	0–10
	α	0.85	0.82	0.84	0.85	0.84	0.84
	PSC case rate (% positive)	16.9	9.6	23.3	17.7	18.0	17.7
	At-risk (top 5%)	9	8	9	9	9	9
	Borderline (top 10%)	8	6	8	8	8	8
Total scores	M	9.49	7.89	11.62	10.01	10.37	10.20
	SD	6.41	5.47	6.51	5.82	6.48	6.26
	Range	0–34	0–29	0–32	0–28	0–30	0–31
	α	0.90	0.87	0.89	0.86	0.89	0.88
	PSC case rate (% positive)	20.5	13.0	29.5	21.3	25.1	24.4
	At-risk (top 5%)	22	18	24	22	23	21
	Borderline (top 10%)	18	16	21	19	19	19

Data not reported for non-binary gender due to the low number of cases.

psychometric properties and validates parent ratings for children aged 3–17 years, in terms of factor structure, reliability and validity. Considered alongside the factor loadings, the three-factor model was supported across preschool and school-age ranges, which is consistent with previous studies which included 4- to 15-year-old children (DiStefano et al., 2017; Murphy et al., 2016).

Overall, using specified evaluation criteria, the PSC-17 for preschool parents was found to have good internal consistency, excellent concurrent validity and good predictive validity. For school parents, the PSC-17 was found to have good internal consistency, good test–retest reliability and excellent concurrent validity. In terms of reliability, the PSC-17 demonstrated strong internal consistency across ages. Test–retest reliability fell just below acceptable levels for the preschool sample according to the specified evaluation criteria; however, this analysis was hampered by a large variation and lengthy time period between testing. Given the rapid child development that occurs during the preschool years, it is not surprising that preschool ICCs were lower than school-age coefficients. The acceptable test–retest reliability for school-age children was higher than other studies which had longer, 6-month retest periods (ICC = 0.55) (Jacobson et al., 2019) and on par with more ideal 8- to 14-day retest period (ICC = 0.85) (Murphy et al., 2016).

In terms of concurrent validity, the PSC-17 correlated strongly with the CBCL across subscales and total scores for both samples. Moreover, ROC analyses revealed the PSC-17 had excellent classification accuracy with very high AUC results across ages. The AUC for the internalising subscale for the preschool sample was borderline acceptable and acceptable for the school sample, which was lower than the AUC of 82% found in the original derivation study (Gardner et al., 1999). That study included a large sample of clinical and primary care presentations which may explain the difference.

In terms of predictive validity, the PSC-17 demonstrated excellent screening accuracy for total scores, externalising and attention subscales using the recommended cut-offs for preschool-age children. Surprisingly the PSC-17 did not perform as strongly in the older age ranges. While excellent specificity was found for all subscales and total scores for school-age children and positive predictive values (PPV) were acceptable for externalising and total scores, only the externalising subscale passed benchmark standards for combined high sensitivity, specificity and PPV. Results may have been hampered by a low number of clinical cases, as is common when recruiting from community samples. Further work is required to investigate sensitivity for school-age children since numerous studies have attained high sensitivity and high specificity for the PSC-17 in this age range (Lavigne et al., 2016).

Normative data

The second aim of this study was to ascertain age and gender normative data for parent-report PSC-17 in Australia for children aged 3–17 years. This study found positive screening rates of 17% for preschool children and 25.3% for school-age children, which is higher than the most recent Australian child and adolescent survey of MH which identified 13.9% of children aged between 4 and 17 years as having a mental disorder (Lawrence et al., 2015) and higher compared to the two previous normative studies from the United States, which were reported as 15% for children of 4–15 years (Gardner et al., 1999) and 11.6% (Murphy et al., 2016).

The higher number of cases may be due to differences in the local population and timing of data collection. It has been established over the past decade that rates of child MH in Australia are high and stable (Sawyer et al., 2018), and Australian-born children are more likely to be positively identified on screening and have mental illness compared to children born elsewhere (Australian Institute of Health and Welfare, 2020). Cut-offs on similar measures are also higher in Australia compared to samples in the United Kingdom and the United States (Kremer et al., 2015).

Elevated scores of child MH may also be due to the effects of the COVID-19 pandemic on Australian children, who experienced a national lockdown, restricted access to schooling and social supports (Sicouri et al., 2023). Research has shown Australian parents had poorer MH and functioning indicating the pandemic had a significant, negative impact on Australian families (Westrupp et al., 2023). While robust systematic comparisons of current child MH symptoms to pre-pandemic data is still emerging, this study's results suggest further investigation is required into the ostensibly high rates of children identified with emerging MH difficulties and is a consideration for the generalisability of the results.

Acceptability

The final aim of this study was to examine acceptability. Results indicated that parents found the PSC-17 to be highly appropriate, easy to understand and important in terms of assessing their child's MH, and there was no difference in levels of acceptability between preschool and school-age parents. The construct of acceptability is associated with the likelihood that screening measures are adopted in future, and thus, these results are a promising indication of parental support for the use of this measure (Glover and Albers, 2007; Kamphaus et al., 2007). High levels of acceptability may be associated with an increased parental awareness of child MH challenges in the aftermath of the pandemic.

Strengths and limitations

Strengths of this study included the large and diverse sample recruited from across Australia and the use of a comprehensive, comparative measure in the form of the CBCL. However, the representativeness of the sample cannot be determined beyond comparative analysis to census data. Moreover, the CBCL is not a gold-standard, criterion measure nor is it a diagnostic measure. Furthermore, recruiting from the community combined with a small sub-sample at the second time point meant that a limited number of clinical cases were identified. This affected the predictive validity analysis; however, low clinical cases are symptomatic of recruiting from a non-clinical population. Common method variance may have also affected results since both assessment measures were questionnaires. Future studies should seek to replicate this study's findings and employ alternate methods of assessment such as comprehensive clinical interviews to assess diagnostic and screening accuracy. Furthermore, future research may investigate the measure's responsiveness and sensitivity to change which are important measurement properties in clinical and research contexts.

Implications

The findings of this study support the PSC-17 as a valid, reliable and acceptable parent measure for preschool- and school-age children within this population. As a freely available, brief measure, it can be adopted widely in a range

of contexts. By spanning the ages of 3–17 years, it enables measurement continuity, i.e., using the same measure at different ages, which is particularly important for longitudinal research, in which children often age out of measures and cross-measure comparisons are difficult. Continuity of measures may also benefit clinical work such as measurement-based care or universal screening conducted at discrete timepoints in a child's life, e.g., at preschool, school entry or high school graduation. In fact, recent surveys have found practitioners perceive brief and free measures as important facilitators to increase the use of measures in practice (Tully et al., 2024). Thus, the PSC-17 may serve as an important instrument for tracking the progress of clinical interventions to improve child MH.

Conclusion

The current study contributes new evidence for the PSC-17 as an appropriate and acceptable measure of child MH, which can be utilised in clinical and community settings for a wide range of ages including preschool-aged children. This study was the first to ascertain normative data for the PSC-17 outside of the United States and for children aged 3 years. By ascertaining normative data for the Australian population and young children, children with emerging MH difficulties can be accurately identified and directed towards further assessment and early intervention.

Acknowledgements

Thanks to Domna Alloush for contributing to data preparation.

Author Contributions

RM designed the study, analysed and interpreted the data and drafted the manuscript.
LT was involved in the conception and the design of the study, reviewed the data and reviewed the manuscript.
MD was involved in the conception and the design of the study, reviewed the data and reviewed the manuscript.
All authors contributed to the development of the research questions; read, reviewed and approved the final manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This research was supported by Growing Minds Australia: A National Trials Strategy to Transform Child and Youth Mental Health Services, and the Growing Minds Australia Clinical Trials Network (Medical Research Future Fund: MRF2006438).

Ethical Approval and Informed Consent statements

Study approval was obtained from the University of Sydney Human Research Ethics Committee (2023/704) on 24 October 2023. All participants provided written informed consent prior to participating.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request and subject to ethical considerations.

ORCID iDs

Rebecca K McLean https://orcid.org/0000-0002-2691-8499
Lucy A Tully https://orcid.org/0000-0002-2066-7609

Supplemental material

Supplemental material for this article is available online.

References

Achenbach TM (1999) The child behavior checklist and related instruments. In: Maruish ME (ed.) *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*, 2nd Edition. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, pp. 429–466.

Australian Institute of Health and Welfare (2020) *Australia's children*. Canberra, ACT, Australia: AIHW.

Aylward GP (1997) Conceptual issues in developmental screening and assessment. *Journal of Developmental and Behavioral Pediatrics: JDBP* 18: 340–349.

Becker-Haimes EM, Tabachnick AR, Last BS, et al. (2020) Evidence base update for brief, free, and accessible youth mental health measures. *Journal of Clinical Child and Adolescent Psychology* 49: 1–17.

Brinley SK, Tully LA, Carl T, et al. (2024) Universal child mental health screening for parents: A systematic review of the evidence. *Prevention Science* 25: 798–812.

Chaffin M, Campbell C, Whitworth DN, et al. (2017) Accuracy of a pediatric behavioral health screener to detect untreated behavioral health problems in primary care settings. *Clinical Pediatrics* 56: 427–434.

D'Souza S, Waldie KE, Peterson ER, et al. (2017) Psychometric properties and normative data for the preschool strengths and difficulties questionnaire in two-year-old children. *Journal of Abnormal Child Psychology* 45: 345–357.

Reyes AL and Langer DA (2018) Assessment and the journal of clinical child and adolescent psychology's evidence base updates series: Evaluating the tools for gathering evidence. *Journal of Clinical Child and Adolescent Psychology* 47: 357–365.

DiStefano C, Liu J and Burgess Y (2017) Investigating the structure of the pediatric symptoms checklist in the preschool setting. *Journal of Psychoeducational Assessment* 35: 494–505.

Frances A (2012) Better safe than sorry. *Australian & New Zealand Journal of Psychiatry* 46: 695–696.

Gao R, Raygoza A, Distefano C, et al. (2022) Assessing measurement equivalence of PSC-17 across teacher and parent respondents. *School Psychology International* 43: 477–495.

Gardner W, Lucas A, Kolko DJ, et al. (2007) Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *Journal of the American Academy of Child and Adolescent Psychiatry* 46: 611–618.

Gardner W, Murphy M, Childs G, et al. (1999) The PSC-17: A brief pediatric symptom checklist with psychosocial problem subscales. *Ambulatory Child Health* 5: 225–236.

Glover TA and Albers CA (2007) Considerations for evaluating universal screening assessments. *Journal of School Psychology* 45: 117–135.

Harrison JR, Vannest KJ and Reynolds CR (2013) Social acceptability of five screening instruments for social, emotional, and behavioral challenges. *Behavioral Disorders* 38: 171–189.

Hawes DJ, Lechowicz M, Roach A, et al. (2021) Capturing the developmental timing of adverse childhood experiences: The adverse life experiences Scale. *The American Psychologist* 76: 253–267.

Hu L-T and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6: 1–55.

Humphrey N and Wigelsworth M (2016) Making the case for universal school-based mental health screening. *Emotional and Behavioural Difficulties* 21: 22–42.

IBM SPSS (2022) Amos. 29.0.0 ed. [Computer Program] Chicago, IL.

Jacobson JH, Pullmann MD, Parker EM, et al. (2019) Measurement based care in child welfare-involved children and youth: Reliability and validity of the PSC-17. *Child Psychiatry and Human Development* 50: 332–345.

Jellinek MS and Murphy JM (1988) Screening for psychosocial disorders in pediatric practice. *American Journal of Diseases of Children (1960)* 142: 1153–1157.

Jureidini J and Raven M (2012) Healthy Kids check: Lack of transparency and misplaced faith in the benefits of screening. *The Australian and New Zealand Journal of Psychiatry* 46: 924–927.

Kamphaus RW, Thorpe JS, Winsor AP, et al. (2007) Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement* 67: 342–356.

Kataoka SH, Zhang L and Wells KB (2002) Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry* 159: 1548–1555.

Kendall PC and Sheldrick RC (2000) Normative data for normative comparisons. *Journal of Consulting and Clinical Psychology* 68: 767–773.

Kessler RC, Amminger GP, Aguilar-Gaxiola S, et al. (2007) Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry* 20: 359–364.

Kremer P, deSilva A, Cleary J, et al. (2015) Normative data for the strengths and difficulties questionnaire for young children in Australia. *Journal of Paediatrics and Child Health* 51: 970–975.

Lavigne JV, Meyers KM and Feldman M (2016) Systematic review: Classification accuracy of behavioral screening measures for use in integrated primary care settings. *Journal of Pediatric Psychology* 41: 1091–1109.

Lawrence D, Johnson S, Hafekost J, et al. (2015) *The mental health of children and adolescents: Report on the second Australian child and adolescent survey of mental health and wellbeing*. August, Department of Health, Canberra, ACT, Australia.

Liu J, Burgess Y, DiStefano C, et al. (2020) Validating the pediatric symptoms checklist–17 in the preschool environment. *Journal of Psychoeducational Assessment* 38: 460–474.

McLean RK, Tully LA, Brinley SK, et al. (2024) Multi-informant universal mental health screening for preschool-aged children by parents and educators: A PRISMA systematic review. *Clinical Child and Family Psychology Review* 28: 39541081.

Mellor D (2005) Normative data for the strengths and difficulties questionnaire in Australia. *Australian Psychologist* 40: 215–222.

Moore SA, Dowdy E, Fleury I, et al. (2021) Comparing informants for mental health screening at the preschool level. *School Psychology Review* 2022; 51: 589–608.

Mulraney M, Coghill D, Bishop C, et al. (2021) A systematic review of the persistence of childhood mental health problems into adulthood. *Neuroscience and Biobehavioral Reviews* 129: 182–205.

Murphy JM, Bergmann P, Chiang C, et al. (2016) The PSC-17: Subscale scores, reliability, and factor structure in a new national sample. *Pediatrics* 138: e20160038.

Oh E, Mathers M, Hiscock H, et al. (2015) Professional help seeking for young children with mental health problems. *Australian Journal of Psychology* 67: 187–195.

Rhodes A (2017) *Child mental health problems: Can parents spot the signs?* 10 October, The Royal Children's Hospital Melbourne, Parkville, VIC, Australia.

Robinson M, Oddy WH, Li J, et al. (2008) Pre- and postnatal influences on preschool mental health: A large-scale cohort study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 49: 1118–1128.

Sawyer MG, Reece CE, Sawyer ACP, et al. (2018) Has the prevalence of child and adolescent mental disorders in Australia changed between 1998 and 2013 to 2014? *Journal of the American Academy of Child and Adolescent Psychiatry* 57: 343–350.

Semansky RM, Koyanagi C and Vandivort-Warren R (2003) Behavioral health screening policies in Medicaid programs nationwide. *Psychiatric Services* 54: 736–739.

Sheldrick R, Henson BS, Merchant S, et al. (2012) The preschool pediatric symptom checklist (PPSC): Development and initial validation of a new social/emotional screening instrument. *Academic Pediatrics* 12: 456–467.

Sicouri G, March S, Pellicano E, et al. (2023) Mental health symptoms in children and adolescents during COVID-19 in Australia. *Australian & New Zealand Journal of Psychiatry* 57: 213–229.

Spencer AE, Baul TD, Sikov J, et al. (2020) The relationship between social risks and the mental health of school-age children in primary care. *Academic Pediatrics* 20: 208–215.

Stoppelbein L, Greening L, Moll G, et al. (2012) Factor analyses of the pediatric symptom checklist-17 with African-American and Caucasian pediatric populations. *Journal of Pediatric Psychology* 37: 348–357.

Tully LA, Kan J, Turnell A, et al. (2024) Use of measures and measurement-based care in child and youth mental health: A survey of Australian practitioners. *Australian Journal of Psychology* 76: 2426662.

Wagner JL, Guilfoyle SM, Rausch J, et al. (2015) Psychometric validation of the pediatric symptom checklist-17 in a pediatric population with epilepsy: A methods study. *Epilepsy & Behavior* 51: 112–116.

Westrupp EM, Bennett C, Berkowitz T, et al. (2023) Child, parent, and family mental health and functioning in Australia during COVID-19: Comparison to pre-pandemic data. *European Child & Adolescent Psychiatry* 32: 317–330.

Youngstrom EA, Van Meter A, Frazier TW, et al. (2017) Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice* 24: 331–363.

Child and Adolescent Psychiatry and Mental Health

RESEARCH

Open Access

Multi-informant reports of preschool mental health: Validation of parent and educator reports and normative data for the preschool Pediatric Symptom Checklist and PSC-17

Rebecca K. McLean[1*], Lucy A. Tully[1] and Mark R. Dadds[1]

Abstract

Background The prevalence of mental health problems and unmet need in preschool-age children highlight the challenge of identifying emerging difficulties using validated measures. Given the lack of existing brief screening measures for preschool-age children, especially multi-informant measures, this study examined two versions of the Pediatric Symptom Checklist (PSC) as reported by parents and educators, the *Preschool Pediatric Symptom Checklist* (PPSC) and the *PSC-17*. In line with Standards for Reporting Diagnostic accuracy studies (STARD) guidelines, this study examined the psychometric properties, scoring thresholds, and acceptability of parent-reported PPSC. It also examined the psychometric properties of educator ratings for both PSC measures and the incremental validity of educator and parent ratings.

Methods Two studies present validation evidence for two mental health measures for use with preschool-age children. Participants were a nationally representative sample of Australian parents ($n = 1,045$; study 1) and a paired sample of parents and educators ($n = 94$ dyads; study 2) of children aged 3–5 years.

Results Results supported the internal consistency, test-retest reliability, concurrent validity of the PPSC. Parents and educators indicated high levels of acceptability of both PSC measures. Results indicated parent-reported PPSC and PSC-17 significantly improved the prediction of clinician-rated functioning scores over and above educator report suggesting incremental validity for multi-informant report. Normative data for the parent-reported PPSC are presented for the first time.

Conclusions This research expands the evidence base for the validity, reliability and acceptability of the parent and educator-report PPSC and PSC-17 measures as utilised with young children. Although further research is required, this research contributes new evidence, including incremental validity and normative data, to increase the clinical utility of both PSC measures.

Keywords Child mental health, Screening, Assessment, Multi-informant, Educator

*Correspondence:
Rebecca K. McLean
rebecca.mclean@sydney.edu.au
[1]School of Psychology, University of Sydney, Camperdown, NSW 2050, Australia

Despite the early onset and prevalence of mental health (MH) problems in preschool children aged under 5 years [1], young children are less likely to access services for their MH concerns compared to adolescents [2]. Prevalence estimates of child MH problems in preschool-aged children vary widely from 7.1 to 26.4% [3–5] and without access to services or treatment, problems can remain stable, persist into adulthood and have long-term negative ramifications [6–8]. Early identification of MH problems can improve outcomes [9]; however, in order to identify children in need of intervention, validated measures suitable for use in the preschool population, using multi-informant reports from primary caregivers, are needed. The *Preschool Pediatric Symptom Checklist* (PPSC; [10]) and *Pediatric Symptom Checklist-17* (PSC-17; [11]) are broad-based child MH measures, but there are significant gaps in the research supporting these measures with parents and educators of preschool-aged children; they require further validation and normative data in order to have greater clinical utility, and be adopted in research and clinical practice. This study presents new validation evidence for the PPSC and PSC-17 for parent and educator reports for preschool-age children and provides normative data for the PPSC for the first time.

Universal screening

The prevalence of problems in preschool-age children highlights a need to identify MH difficulties early so that children can receive intervention. One tool for early identification is a population-based approach called universal screening, which is also known as systematic monitoring, routine surveillance, or wellbeing check-ins. Screening involves the use of standardised measures, such as the Patient Health Questionnaire [12] and Spence Children's Anxiety Scale [13] to identify children who may be at risk of MH problems. The measures are rated by single or multiple informants, and generally quick and inexpensive to implement [14]. Screening is particularly important in the early years when MH problems are less entrenched and severe [15].

Screening measures or check-ins can be useful for parents, health professionals, and educators. Completion of wellbeing check-ins may lead to an opportunity to discuss child MH, since research suggests few parents or educators will raise concerns with each other spontaneously, and instead will wait for other caregivers to alert them to potential concerns [16, 17]. This emphasises the importance of developing multi-informant tools that are considered acceptable by parents and educators.

Multi-informant screening utilising parents and educators

Whilst reports between raters are often not highly consistent, multi-informant assessment is a hallmark of developmental psychology and essential for early identification [18, 19]. Multi-informant measures are needed to identify children at risk of MH problems; however, two recent systematic reviews have highlighted there are few measures validated for use with preschool-aged children either with single or multiple raters [20, 21]. In order to allow multi-informant reports, there is a need, therefore, to establish psychometrically sound measures for young children that can be used with parents and educators.

Early childhood education and care (ECEC) educators' perspectives of child MH are valuable, as they observe children in naturalistic play for many hours and can compare child MH with a referent group of peers [22, 23]. Whilst medical professionals are increasingly using screening measures in their practice [24], educators often lack MH training and knowledge around measure use [25]. Yet, educator reports are one of the primary pathways of identification of MH problems in young children [26] and educators are a primary source of advice when parents are help-seeking for their child [27]. There is, therefore, a need to provide parents and educators with measures that will help them identify children at risk of MH difficulty.

Whilst adding educator perspectives to parent-report measures has the potential to increase the validity of screening, it can also raise the possibility of discrepant informant ratings [18]. Although discrepancies between informants may appear disadvantageous, they can in fact offer important, domain-specific information about a child's behaviour informed by the informant's context [28]. When multiple perspectives about a child are captured from various contexts like the home and preschool, the ideal model of child assessment is more likely to be obtained [29]. Moreover, incremental validity is important for multi-informant measures as it provides an indication of how much each informant's perspective contributes to a specified outcome and whether additional perspectives are justified. However, incremental validity of multi-informant measures for preschool children has not been well researched. A recent systematic review identified only one study investigating incremental validity of two measures; a significant improvement in identification of child MH problems was found when educator ratings were added to parent ratings for the PSC-17 and Behavior Assessment System for Children-3 Behavioral and Emotional Screening System [21]. The lack of research in this area reveals a need to understand the incremental value of collecting data from multiple caregivers.

The Pediatric Symptom Checklists

Whilst there are several MH screening measures, which have been identified in previous reviews of the literature, only a few have been validated for use with preschool-age children and these measures have varying psychometric strength [20, 30]. Psychometric properties such as internal consistency, test-retest reliability, inter-rater reliability, and predictive validity are important for ensuring measures are free from measurement error and are suitable for use with the population in which they were tested. Two *Pediatric Symptom Checklists* (PSCs) have been recommended for use with preschool populations: the PSC-17 [11] and PPSC [10]. Both are brief, free in digital and paper-based formats, and easy to score. They produce total scores and subscales based on internalising, externalising and attention problems. The PSC-17, which is widely used, was developed for older children, but has recently been validated for use with children aged 3–5 years [31]. However, investigations of educator report for the PSC-17 are limited [32–34]. Previous studies have investigated the factor structure of educator-rated PSC-17, inter-rater reliability and incremental validity in the preschool environment [32, 34, 35], however, other facets of reliability and validity of educator report with a paired sample of parents and educators have not been investigated for the PSC-17. Further, the convergence of educator-report with diagnostic clinical interviews has not been investigated for this measure. The 18-item PPSC is a related measure, adapted from the PSC-17 and specially developed for preschool-aged children 18 months–5.5 years, with developmentally appropriate item wording and a fourth subscale called "Parenting Challenges", which assesses parent difficulties across six items. However, the developers of this scale caution against interpreting any of the subscales, due to concerns over reliability, and instead suggest use of the total score only. Whilst previous research on the PPSC features diverse samples [36, 37], there is no normative data and the measure has not yet been examined with educators, limiting its clinical utility. The PPSC has strong internal consistency, Cronbach's alpha 0.80–0.92, and good test-retest reliability (intraclass correlation coefficients [ICC] = 0.75) after four weeks [10, 38]. Classification accuracy or predictive validity is important for screening measures; the PPSC has high sensitivity (0.87 – 1.00) when tested against the Child Behaviour Checklist (CBCL) and the Ages & Stages Questionnaires [10]. Recent research reported moderate predictive validity of sensitivity 0.75, specificity 0.77, Positive Predictive Value (PPV) 0.60, Negative Predictive Value (NPV) 0.99 for the PPSC compared to International Classification of Disease scores after 1 year [39].

Clinical thresholds or cut-off points in screening measures determine the classification of children and their level of risk for MH problems [40]. Determining appropriate clinical thresholds is key to early identification. If thresholds are too low, healthcare services risk being overwhelmed and parents unnecessarily concerned; too high and at-risk children may be overlooked for intervention [40]. It is, therefore, important for researchers to determine appropriate thresholds and normative data based on the prevalence of child psychopathology and the population of interest. No normative data have been established for the PPSC. Whilst the PSC-17 has been the focus of a number of studies [31], more limited research on the psychometric properties of the PPSC warrants further attention. Thus, this study includes an examination of both PSC measures that are appropriate for use with preschool populations.

Acceptability

Acceptability in terms of whether a measure is perceived as easy to comprehend, helpful or useful can affect implementation of screening [15, 41]. Acceptability may increase uptake and implementation of screening programs and therefore, it is an important aspect to investigate if measures are to be adopted as part of assessment and screening practices [42]. The acceptability of the PPSC, as reported by any user, has not been explored. Investigating whether educators find the PSC measures acceptable is important if these measures are to be adopted in ECEC services.

The current study

No research has investigated the PPSC with educators, little research has investigated incremental validity for either PSC measure and no studies have investigated acceptability of the PPSC. Thus, the overall aims were to examine validity and establish normative data for parent-reported PPSC, examine the validity of educator-reported PPSC and PSC-17, and examine acceptability of parent- and educator-reported PPSC.

Two studies were conducted with two cohorts. The first study examined the validity and acceptability of parent-reported PPSC, and the second study examined the validity of educator-reported PPSC and PSC-17, and incremental validity. Analyses were conducted using SPSS 29.0.1.0 [43].

Study 1

Study 1 aimed to: (1) examine the psychometric properties of parent-reported PPSC, in terms of internal consistency, test-retest reliability, concurrent and predictive validity, and to also examine scoring thresholds; (2) establish normative data for PPSC in Australia; and (3) examine the acceptability of this measure.

Method

Procedures

Whilst there are no set conventions for sample sizes for normative data, this study aimed to recruit a large, representative sample from the Australian population using an online research panel. Participants were recruited through an Australian, online research panel and sampled to represent families based on key socio-demographic characteristics from national census data (i.e., household income, marital status, and residential location) using segmented quotas. After sampling, parent demographic details including household income, marital status, and residential location (metropolitan, regional/rural) were compared to census data to ensure key characteristics of the sample aligned with the broader population [44–46]. Adult parents or caregivers of a child aged 3–5 years, who lived in Australia and understood English completed baseline measures (N = 1,045), which were counterbalanced in order of completion. Participants who completed baseline questionnaires were invited to participate in a second assessment. To evaluate test-retest reliability, concurrent and predictive validity, a sub-sample of participants (n = 51) completed a second set of questionnaires on average 35 days after baseline. Participant flow is presented in Fig. 1. See supplementary material for additional details about recruitment procedures.

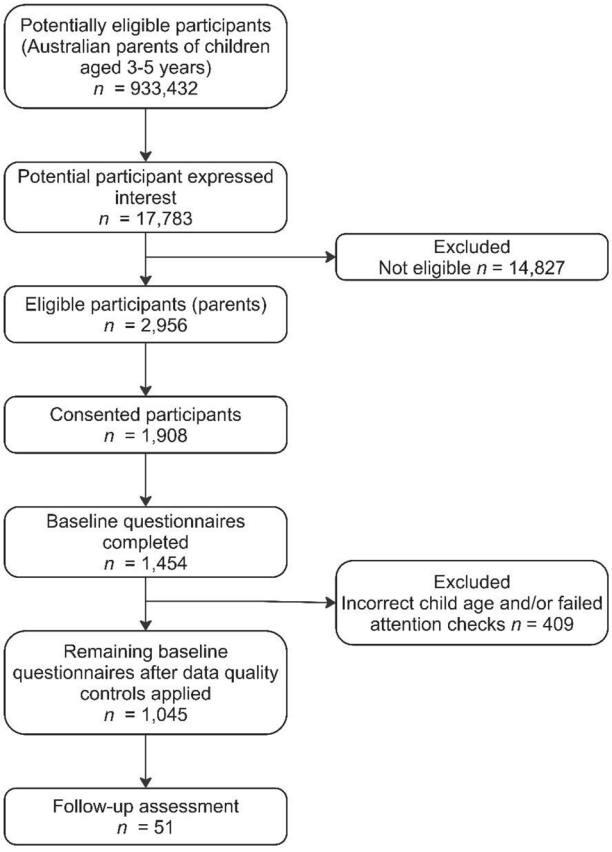


Fig. 1 Flow of participants through Study 1

Participants

Mean child age was 3.87 years (SD = 0.78) and 46.89% were female. Child ethnicity was predominantly Caucasian, but diverse (See supplementary material for additional details about recruitment procedures). Parents were aged 36.30 years on average (SD = 8.10); 66.6% of parents identified as female; and 84.1% of parents were married or in *de facto* relationships.

Measures

In addition to the PPSC, parents' views on the acceptability of the PPSC were assessed using four items adapted from a parent-report measure developed by Hawes et al [47]. Responses were collected on a 5-point Likert response scale from *not at all/strongly disagree (1)* to *very/strongly agree (5)* . Items index confidence in ratings ("Based on your knowledge of your child, how confident are you in the responses/ratings you have given today as part of the Checklist?"), the appropriateness ("How appropriate did you find the wording of the questions, i.e., were the questions appropriate for your child's age or developmental stage?" and "The Checklist asks important questions about my child's mental health") and clarity of item wording ("How clear did you find the wording of the questions, i.e., were they easy to understand?").

The CBCL for 1.5-5 years [48] was used as a criterion measure at the second assessment point to establish concurrent and predictive validity with PPSC. The CBCL is a comprehensive, 99-item measure of child MH. Cases were classified as borderline, clinical, or at-risk including both borderline and clinical cases. The CBCL has strong psychometric properties and is the most common criterion measure in classification accuracy studies [30].

Analytic plan

To examine the first aim, Cronbach's alpha for internal consistency and ICC for test-retest reliability were assessed. ICC estimates and their 95% confident intervals were calculated based on a single rating, absolute-agreement, 2-way mixed-effects model. To assess concurrent validity with the CBCL, Pearson correlations were calculated. PPSC scores were classified into at-risk categories using the recommended scoring threshold. Receiver operator characteristic (ROC) analysis was used to evaluate concurrent validity based on categorical risk classifications and to examine recommended thresholds. Low cell counts meant assumptions were violated for Chi-Square tests. Fisher's Exact Test was not significant thus we were unable to analyse the predictive validity of the PPSC compared to the CBCL clinical or at-risk cases as intended. Psychometric properties were evaluated

according to the rubric developed by Youngstrom et al [49] and De Los Reyes and Langer [50].

To address the second aim, normative data was determined by reporting scoring cut-offs for the 90th (borderline) and 95th percentile (at-risk) for overall age and gender ($N = 1{,}045$). To address the third aim of acceptability, frequency distributions were examined ($n = 51$).

Since the online survey required input for all items, there was no missing data.

Table 1 Mean scores and banding for PPSC subscales and total scores

		3–5 years		
		(*N* = 1,045)	**Male (*n* = 554)**	**Female (*n* = 490)**
Internalising	M	3.06	3.28	2.81
	SD	2.47	2.56	2.35
	Range	0–12	0–12	0–12
	Possible range	0–12		
	α	0.777	0.790	0.757
	At-risk (top 5%)	8	8	7
	Borderline (top 10%)	7	7	6
Externalising	M	1.30	1.56	1.01
	SD	1.65	1.82	1.39
	Range	0–8		
	Possible range	0–8	0–8	0–8
	α	0.776	0.805	0.705
	At-risk (top 5%)	4	5	4
	Borderline (top 10%)	4	4	3
Attention	M	1.95	2.21	1.67
	SD	1.72	1.81	1.57
	Range	0–6	0–6	0–6
	Possible range	0–6		
	α	0.808	0.825	0.771
	At-risk (top 5%)	6	6	5
	Borderline (top 10%)	5	5	4
Parenting challenges	M	3.24	3.55	2.88
	SD	2.91	3.02	2.75
	Range	0–12	0–12	0–12
	Possible range	0–12		
	α	0.834	0.844	0.816
	At-risk (top 5%)	9	10	9
	Borderline (top 10%)	8	8	7
Total scores	M	9.22	10.21	8.10
	SD	6.81	7.28	6.07
	Range	0–36	0–36	0–36
	Possible range	0–36		
	α	0.900	0.911	0.877
	At-risk (top 5%)	23	25	20
	Borderline (top 10%)	18.4	20	16.9
	PPSC case rate (% positive)	47.4	53.1	40.8

Data not reported for one child identified as non-binary gender

Results

Reliability and validity

Excellent internal consistency (α = 0.90) and adequate test-retest reliability was found for PPSC total scores (ICC = 0.78, CI 0.65–0.87).

PPSC total scores strongly and positively correlated with CBCL total scores at the bivariate level demonstrating concurrent validity, $r = 0.86$, $p < 0.001$, $n = 51$.

ROC analysis examined the association between PPSC and CBCL at-risk case classifications. The area under curve (AUC) for PPSC total scores was 0.89 with a standard error of 0.09. This was significant at $p < 0.001$ (CI 0.72–1.06).

ROC analysis identified an alternative clinical threshold of 19.5 for the PPSC, which maximised sensitivity and specificity. For ease of scoring, the alternate score was rounded down and analysed at 19, which identified 10.0% ($n = 104$) of the sample as at-risk. Chi-square tests of independence indicated an adequate number of cell counts using the alternate threshold. An alternate clinical threshold using a criterion of CBCL clinical cases produced sensitivity of 83.3, specificity 95.6, PPV 71.4, NPV 97.7, and AUC 0.89 (standard error 0.09), $p < 0.001$.

Normative data

The mean, standard deviation and ranges for total scores and subscales were calculated. Using the recommended scoring threshold, 47.4% of the sample ($n = 495$) were identified as at-risk. Table 1 presents scores for the highest-ranking cases by age and gender.

Acceptability

In response to questions of acceptability, most parents agreed or strongly agreed the PPSC asked important questions about their child's MH (86.3%), found the PPSC to be very or quite appropriate (86.3%, $n = 44$) and very or quite clear (96.1%). Most parents were very or quite confident in the ratings they provided (96.1%).

Study 2

Having produced new validation evidence for parent-reported PPSC in study 1, study 2 sought to extend measure utility by validating educator-reported PPSC and PSC-17 and to examine the incremental value of multi-informant ratings by recruiting parents and educators.

The aims of study 2 were to: (1) examine the psychometric properties of educator ratings of the PPSC and PSC-17 in terms of internal consistency, test-retest reliability, inter-rater reliability, predictive validity,

concurrent validity and convergence with diagnostic clinical interviews; (2) examine the incremental validity of parent- and educator-report PPSC and PSC-17 scores using criterion measures including the parent-report CBCL and clinician-rated Children's Global Assessment Scale (CGAS); and (3) examine acceptability of the measures amongst educators.

Method

Procedures

An *a priori* statistical power calculation to estimate an appropriate sample for analysing measurement stability and validity was performed using G*Power 3.1.9.7 [51]. With alpha of 0.05, power 0.80, and an effect size of 0.12, the calculated sample size was 93. Participants were recruited via ECECs and social media advertisements. Adult parents or caregivers of a child aged 3–5 years, who lived in Australia, and understood English were asked for permission to contact their child's educator, as part of the consent process. Eligible educators were adult educators of children aged 3–5 years, who worked in ECEC, had internet access, understood English and knew the child well enough to complete questionnaires about their MH. The final paired sample of parents and educators who completed baseline questionnaires was 94.

Participants completed measures which were counterbalanced in order of completion at each timepoint. Parents completed baseline, 3-month (3mFU) and 6-month follow-up (6mFU) questionnaires. Within two weeks of baseline, parents completed one clinical interview (2wFU), and educators completed baseline questionnaires. Within two weeks of educator baseline, a subsample of educators ($n = 31$) completed a second set of questionnaires to test test-retest reliability. Educators who completed baseline questionnaires were invited to participate in the second assessment until the target recruitment for the sub-sample was reached. Participant flow is presented in Fig. 2.

Participants

Children were aged 3.65 years on average ($SD = 0.67$) and 38.3% female. Child ethnicity was predominantly Caucasian (74.5%), followed by 16% other ethnicities. Educators were predominantly female (97.9%) with a mean of 13.13 years ECEC work experience ($SD = 8.89$). Most educators indicated that they were *very* or *quite* familiar with the child (89.4%, $n = 84$) and had known the child between 2 and 54 months ($M = 13$ months). See supplementary material for additional participant details.

Measures

Educators' views on the acceptability of measures were assessed at baseline using the same measure as reported in Study 1.

Parents completed the CBCL at baseline, 3mFU and 6mFU. The CBCL at baseline and 3mFU was used in this study as a criterion measure to investigate convergent and predictive validity, and as an outcome measure for incremental validity.

Clinicians, who were blind to parents and educators' questionnaires responses, conducted clinical interviews with parents using the *Diagnostic Interview Schedule for Children, Adolescents and Parents* (DISCAP-V; [52]) to assess child psychiatric disorder and the *Children's Global Assessment Scale* (CGAS; [53]) to assess global functioning. The DISCAP-V is a semi-structured interview guide corresponding to common childhood behavioural and affective diagnostic categories. Clinicians rated the severity of the diagnostic presentation (0–6). Those with a severity rating 4 and above were considered to meet criteria for the specified disorder. The CGAS is an outcome measure of a child's general functioning rated on a continuum (0–100); higher scores indicate higher levels of health and superior functioning, while lower scores indicate illness and gross impairment. Parents were blind to clinician ratings.

Clinicians were clinical psychology interns at a university research clinic, registered general and clinical psychologists, all under the supervision of senior clinical psychologists. Inter-rater reliability for a subset of clinical interviews ($n = 24$) was assessed using a secondary team of clinicians, who were blind to the primary clinician's ratings. Diagnostic ratings correlated highly ($r = 0.79$, $p < 0.001$). CGAS scores moderately correlated ($r = 0.57$, $p = 0.004$). There were no significant differences in ratings.

Educators' familiarity with the child was assessed using five items such as how frequently they care for the child, how long they have known the child, level of familiarity and strength of relationship with the child. Responses were collected on a 5-point Likert response scale from *not at all/strongly disagree (1)* [] to *very/strongly agree (5)* [].

Parent-report PPSC was completed at baseline, 3mFU and 6mFU. One item differed for educator-report PPSC: "Is it hard to take the child out in public?" with an additional response option, *Not applicable*. Scoring for this response option was coded as 0. Items were summed to produce a total score. Educators completed this measure at baseline and 2wFU.

The PSC-17 was the same for parents and educators. Parents completed this measure at baseline, 3mFU and 6mFU and educators at baseline and 2wFU.

Analytic plan

The following sub-samples of participants were used to analyse the data: baseline educator ($n = 94$), educator 2wFU ($n = 31$), baseline parent ($n = 178$), parent 2wFU

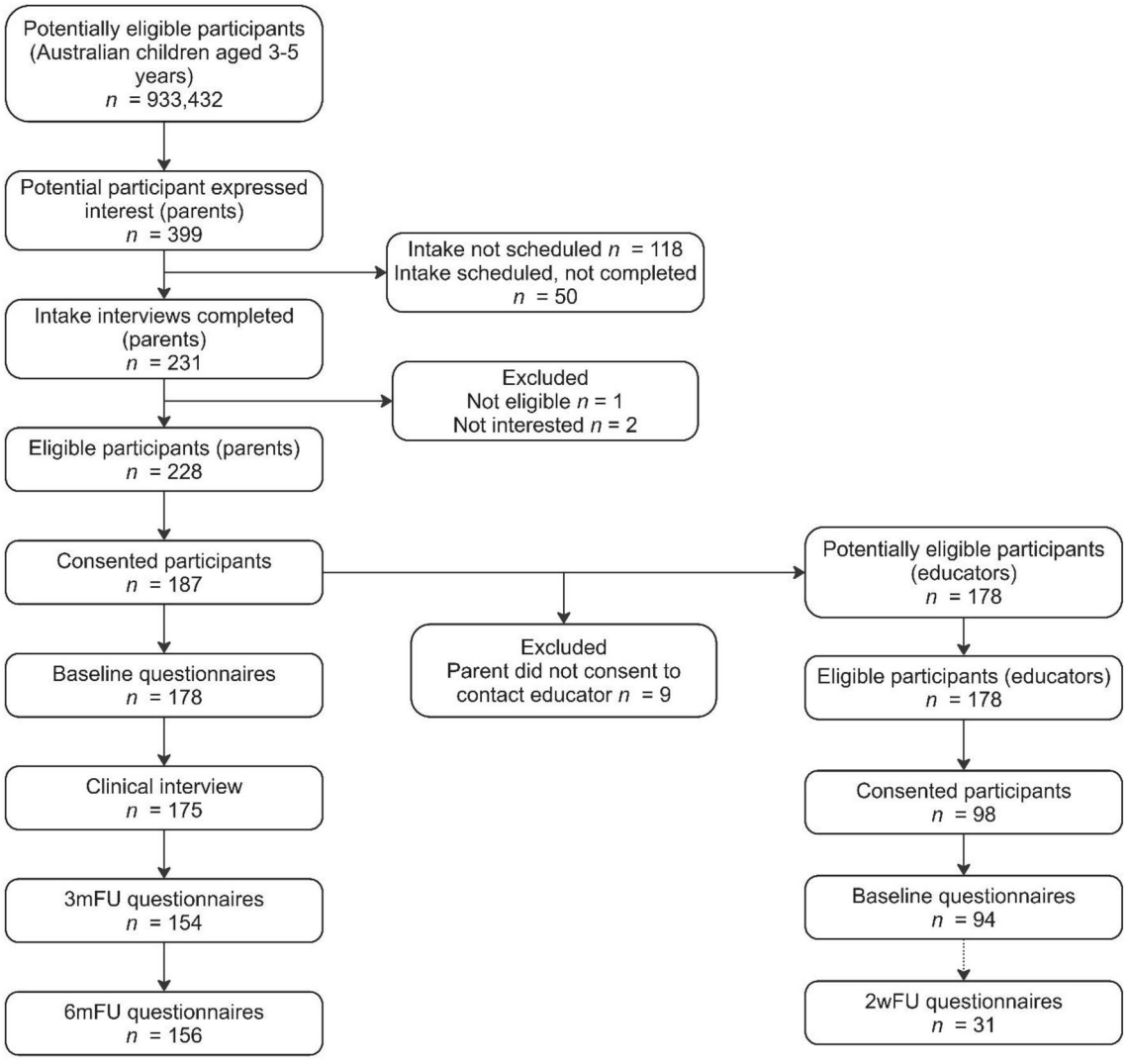


Fig. 2 Flow of participants through Study 2

(*n* = 175, and parent 3mFU (*n* = 154). To examine the first aim, Cronbach's alpha was calculated to assess reliability using baseline educator assessment, ICCs were assessed for test-retest (baseline educator and 2wFU) and inter-rater reliability, and Pearson's correlations for convergent validity with CBCL total scores (baseline parent and educator). ICC estimates and 95% confident intervals were calculated based on a single rating, absolute-agreement, 2-way mixed-effects model. To assess inter-rater agreement, Cohen's κ was calculated using case classifications (baseline parent and educator). The two PSC measures were classified using recommended thresholds. To test concurrent validity with the DISCAP, a series of binary logistic regressions were conducted (baseline educator and parent 2wFU). To assess predictive validity, first cell counts for Chi-Square tests and Fisher's Exact Test were checked to be valid. Baseline educator PPSC and PSC-17 ratings were then compared to parent-rated CBCL clinical cases at 3mFU. Psychometric properties were evaluated according to the same rubric as utilised in Study 1.

To address the second aim, a series of hierarchical multiple regressions were conducted to assess incremental validity (baseline educator, baseline parent and parent 2wFU). To address the third aim of acceptability, frequency distributions were examined for educators at baseline.

Since the online survey required input for all items, there was no missing data for baseline measures for parents and educators, and educator 2wFU. Listwise deletion was used for two cases where parents did not attend clinical interviews. Pairwise deletion was used for parent-reported data missing at 3mFU (*n* = 13).

Results

Reliability and validity

Internal consistency was excellent for baseline educator-reported PPSC (α = 0.91) and good for PSC-17 total scores (α = 0.88). Test-retest reliability for educator-reported PPSC was excellent (ICC = 0.91; CI 0.82–0.96, $p < 0.001$) and good for the PSC-17 (ICC = 0.80; CI 0.66–0.89, $p < 0.001$) for baseline and 2wFU assessments. Inter-rater reliability between baseline parent and baseline educator ratings was small for the PPSC (ICC = 0.33; CI 0.14–0.49, $p < 0.001$) and PSC-17 (ICC = 0.34, CI 0.15–0.50, $p < 0.001$).

Convergent validity was weak and insignificant for baseline educator-reported PPSC and baseline parent-reported CBCL total scores, $r = 0.19$, $p = 0.068$, and baseline educator-reported PSC-17 total scores, $r = 0.19$, $p = 0.067$. Cohen's κ was run to determine if there was agreement between at-risk baseline educator-reported PPSC and baseline parent-reported CBCL at-risk cases, κ = 0.15, CI − 0.05–0.36, $p = 0.123$. When at-risk baseline educator-reported PPSC cases were compared to baseline parent-reported CBCL clinical cases, there was significant, though low, agreement, κ = 0.19, CI − 0.01–0.40, $p = 0.040$. At-risk baseline educator-reported PSC-17 cases compared to baseline parent-reported CBCL at-risk cases found κ = 0.11, CI − 0.11–0.33, $p = 0.275$, and to baseline parent-reported CBCL clinical cases κ = 0.09, CI − 0.14–0.33, $p = 0.360$.

To test concurrent validity, a series of univariate binary logistic regressions were performed with baseline educator-reported PPSC and PSC-17 total scores, and at-risk classifications as the predictors and any type of DISCAP diagnosis as the outcome ($n = 92$). None of the models were significant as presented in Table 2. Neither PSC significantly contributed towards diagnosis.

Predictive validity

The predictive validity of the educator-reported measures was tested against the CBCL using Chi-Square tests. Compared to parent-reported CBCL clinical cases at 3mFU, baseline educator-reported PPSC scores produced sensitivity 24.0%, specificity 89.3%, PPV 50.0 and NPV 72.46; and PSC-17 scores produced sensitivity 36.4%, specificity 88.6%, PPV 33.33 and NPV 89.86.

Classification accuracy as indicated by AUC was moderate for the PPSC (AUC = 0.65, CI 0.51–0.79, $p = 0.037$) and PSC-17 (AUC = 0.63, CI 0.49–0.77, $p = 0.061$).

Table 2 Tests of convergence for educator-report PSC measures with DISCAP diagnoses

Predictor variables Reference: DISCAP diagnosis $n = 92$	χ2, significance	Nagelkerke R^2	Wald
PPSC total scores	1.77, $p = 0.183$	0.03	1.84, $p = 0.175$
PPSC at-risk classifications	1.20, $p = 0.273$	0.02	1.22, $p = 0.269$.
PSC-17 total scores	1.19, $p = 0.276$	0.02	1.21, $p = 0.271$
PSC-17 at-risk classifications	1.26, $p = 0.268$	0.02	1.31, $p = 0.253$.

Coding for case classifications: 0 = not at-risk, 1 = at-risk. Coding for DISCAP diagnosis: 0 = no diagnosis, 1 = one or more diagnoses

Incremental validity

To test incremental validity, hierarchical multiple regressions were conducted, with three blocks of variables. First, child age and gender were entered as predictors with clinician-rated CGAS as the outcome variable, neither were significant, F (2) = 1.15, $p = 0.321$, $R^2 = 0.02$. In the second model, baseline parent-reported PPSC total scores were added and showed significant improvement on the first model F (3) = 14.19, $p < 0.001$, $R^2 = 0.33$. In the third model, baseline educator-reported PPSC total score was added and was significant F (4) = 11.77, $p < 0.001$. It explained an additional 2.5% of variance, however this was not a significant change, indicating no increase in incremental validity when educator reports were included.

When the order of entry was reversed and baseline educator-report PPSC scores were entered in the second model and baseline parent-report scores in the third model, both models were found to significantly improve prediction of clinician-rated CGAS scores after controlling for age and gender indicating an increase in incremental validity when parent reports were included. Hierarchical regressions were also performed with parent-reported CBCL as the outcome variable. Overall, models using the PPSC explained 35.1% in variance of CGAS scores and 50% of CBCL total scores. Table 3 presents a summary of results.

Analysis was repeated with the PSC-17 and clinician-rated CGAS as the outcome variable. After controlling for age and gender, baseline parent-reported PSC-17 was a significant predictor, F (3) = 4.75, $p = 0.004$, and explained 13.9% of variance. The third model was also significant, F (4) = 4.58, $p = 0.002$ and explained an additional 3.5% variance, however this was not a significant change in incremental validity when educator reports were included.

When the order of entry was reversed, both models were found to significantly improve prediction of clinician-rated CGAS scores. After controlling for age and gender, baseline educator-report PSC-17 explained 11.3% of variance, F (3) = 3.73, $p = 0.014$. The third model was also significant F (4) = 4.58, $p = 0.002$ and explained an additional 6.1% of variance, $p = 0.013$ indicating a significant increase in incremental validity when baseline parent reports were included. Overall, models using the

Table 3 Summary of hierarchical regression models for educator-report PPSC and parent-report PPSC total scores as predictors

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CGAS									
Block 1:	0.03	0.03	1.15	2	0.321				
Age						−0.21	0.14	−0.16	0.134
Gender						−0.72	2.23	−0.03	0.748
Block 2:	0.33	0.30**	14.19	3	< 0.001**				
Age						−0.17	0.12	−0.13	0.144
Gender						-1.77	1.87	−0.08	0.347
Parent-reported PPSC						-1.12	0.18	−0.55	< 0.001**
Block 3:	0.35	0.03	11.77	4	< 0.001**				
Age						−0.20	0.11	−0.16	0.083
Gender						-1.97	1.85	−0.09	0.289
Parent-reported PPSC						−0.99	0.19	−0.49	< 0.001**
Educator-reported PPSC						−0.26	0.14	−0.17	0.070
CGAS									
Block 1:	0.03	0.03	1.15	2	0.321				
Age						−0.21	0.14	−0.16	0.134
Gender						−0.72	2.23	−0.03	0.748
Block 2:	0.15	0.12**	5.06	3	0.003*				
Age						−0.26	0.13	−0.20	0.045
Gender						-1.39	2.10	−0.07	0.512
Educator-reported PPSC						−0.54	0.15	−0.35	< 0.001**
Block 3:	0.35	0.20**	11.77	4	< 0.001**				
Age						−0.20	0.11	−0.16	0.083
Gender						-1.97	1.85	−0.09	0.289
Educator-reported PPSC						−0.26	0.14	−0.17	0.070
Parent-reported PPSC						−0.99	0.19	−0.49	< 0.001**
CBCL									
Block 1:	0.01	0.01	0.61	2	0.546				
Age						0.10	0.28	0.04	0.730
Gender						4.92	4.52	0.12	0.279
Block 2:	0.49	0.48**	29.15	3	< 0.001**				
Age						−0.002	0.20	−0.001	0.994
Gender						7.62	3.27	0.18	0.022*
Parent-reported PPSC						2.88	0.31	0.70	< 0.001**
Block 3:	0.50	0.003	21.89	4	< 0.001**				
Age						−0.02	0.20	−0.01	0.904
Gender						7.47	3.28	0.17	0.025*
Parent-reported PPSC						2.97	0.34	0.72	< 0.001**
Educator-reported PPSC						−0.19	0.25	−0.06	0.460
CBCL									
Block 1:	0.01	0.01	0.61	2	0.546				
Age						0.10	0.28	0.04	0.730
Gender						4.92	4.52	0.12	0.279
Block 2:	0.05	0.04*	1.73	3	0.167				
Age						0.16	0.28	0.06	0.555
Gender						5.71	4.47	0.13	0.204
Educator-reported PPSC						0.64	0.32	0.21	0.051*
Block 3:	0.50	0.44**	21.89	4	< 0.001**				
Age						−0.02	0.20	−0.01	0.904
Gender						7.47	3.28	0.17	0.025*
Educator-reported PPSC						−0.19	0.25	−0.06	0.460
Parent-reported PPSC						2.97	0.34	0.72	< 0.001**

Table presents a summary of hierarchical regression models for baseline educator-report PPSC and baseline parent-report PPSC total scores as predictor variables with CGAS and CBCL total scores as outcome variables ($n = 92$)

*Significant at $p < 0.05$, ** $p < 0.01$

PSC-17 explained 17.4% in variance of CGAS scores and 41% of CBCL total scores. Table 4 presents a summary of results.

Acceptability

In response to questions of acceptability, educators found the PSC measures highly acceptable. Educators agreed or strongly agreed the measures asked important questions about the child's MH (PPSC 78.7%, $n = 74$; PSC-17 87.2%, $n = 82$), found the measures to be very or quite appropriate (PPSC 84%; PSC-17 86.1%) and very or quite clear (PPSC 94.7%; PSC-17 93.6%). Educators were very or quite confident in the ratings they provided (PPSC 90.4%; PSC-17 93.6%).

Paired sample t-tests showed there was no significant difference between educator ratings of acceptability between the PPSC and PSC-17 in terms of appropriateness or clarity. In terms of whether the measures asked important questions, educators rated the PSC-17 significantly higher on average ($M = 3.15$, SD 0.69) than the PPSC ($M = 2.98$, SD 0.73), $t(93) = -3.05$, $p = 0.003$.

General discussion

With an aim of examining the psychometric properties of parent-reported PPSC, the first study contributed new reliability and validity evidence for the measure, finding excellent internal consistency, adequate test-retest reliability and strong concurrent validity with the CBCL total scores and case classifications. Given the high proportion of children identified as at risk in this sample using the original threshold, alongside consideration of new normative data addressing the second aim, an alternative threshold was identified as 19, which is substantially higher than the original cut-off of 9. High sensitivity and specificity using alternative thresholds surpassed benchmark standards for developmental screening measures [30, 54].

This study was the first to explore acceptability of the parent-rated PPSC—the study's third aim. Acceptability of the PPSC by parents was very high. More than 85% of parents rated the PPSC as appropriate on all indices. Our results align with previous findings with older children across health and education settings, in which parents perceived universal MH screening as appropriate and helpful [41].

The second study complemented the first by examining the psychometric properties of educator-reported PPSC and PSC-17. Very high internal consistency and test-retest reliability was found for both measures. Unsurprisingly, inter-rater reliability between parent and educator ratings was small, but significant for both measures, which corresponds with previous findings in which there is low correspondence between informants when they observe child behaviour in different settings [18]. Specifically for the PSC-17, our findings whilst small (ICC = 0.34, CI. 15-0.50, $p < 0.001$) were higher than found in a previous study on the measure, which reported a bivariate correlation of $r = 0.26$, $p < 0.001$ [34]. This was the first study to test inter-rater reliabilities for parent and educator reports on the PPSC so there are no comparisons to make in the literature. This study's results are comparable to other studies of agreement between parent and teacher reporters on screening measures of anxiety (r 0.21 – 0.32, $p < 0.01$) [55], although cross-informant agreement does vary widely across measures [21]. For analyses in which the outcome variable was parent-reported CBCL or clinician-rated diagnosis (concurrent, convergent, predictive and incremental validity), educator ratings did not perform well overall, although educator classification accuracy was moderate, as indicated by AUC analysis and educator ratings for both PSC measures had high specificity. These results may reflect informant discrepancies, possible bias of the parent-report criterion measure or criterion contamination.

Addressing the second aim of incremental validity, results indicated that incremental validity of parent report in addition to educator report was demonstrated for both PSC measures when the outcome was parent-reported CBCL scores. Educator reports did not add significant variance for the parent-reported outcome. Educator-reported PSC measures significantly improved the prediction of clinician-rated functioning scores, only when educator report was entered before report. In contexts where educators are the first gate, such as an ECEC setting, parents add significantly more variance to the model, which contrasts with previous research that found educators added significantly information on top of parent ratings for the PSC-17 [34]. Interestingly, the PSC-17 did not explain as much variance in these models, suggesting the PPSC's items assessing parenting challenges may more closely align with impairment and functioning. These findings signal the importance of gathering parent ratings on child MH in addition to educator reports. Despite challenges facing time-poor caregivers, this research highlights the value of multi-informant assessment in young children.

To address the final aim, educators rated both measures as highly acceptable. Interestingly, educators rated the PSC-17 more highly than the PPSC in terms of whether it asked important questions; a difference not reflected in parent ratings. Despite concerns that some PSC-17 items are not well-suited to this age group, results from both studies indicate the PSC-17 was perceived by parents and educators as being age-appropriate and highly acceptable. These findings align with previous research, which found parents and educators find screening programs acceptable, and parents of preschool-age children rate the PSC-17 as an acceptable screening measure [21,

Table 4 Summary of hierarchical regression models for educator-report PSC-17 and parent-report PSC-17 total scores as predictors

	R^2	ΔR^2	F	df	Sig.	B	S.E.	β	Sig.
CGAS									
Block 1:	0.03	0.03	1.15	2	0.321				
Age						−0.21	0.14	−0.16	0.134
Gender						−0.72	2.23	−0.03	0.748
Block 2:	0.14	0.11**	4.75	3	0.004*				
Age						−0.15	0.13	−0.12	0.254
Gender						-1.97	2.14	−0.09	0.358
Parent-reported PSC-17						−0.85	0.25	−0.35	< 0.001**
Block 3:	0.17	0.03	4.58	4	0.002*				
Age						−0.19	0.13	−0.15	0.146
Gender						-2.41	2.12	−0.12	0.258
Parent-reported PSC-17						−0.67	0.26	−0.27	0.013*
Educator-reported PSC-17						−0.36	0.19	−0.20	0.060
CGAS									
Block 1:	0.03	0.03	1.15	2	0.321				
Age						−0.21	0.14	−0.16	0.134
Gender						−0.72	2.23	−0.03	0.748
Block 2:	0.11	0.09*	3.73	3	0.014*				
Age						−0.25	0.13	−0.19	0.062
Gender						-1.77	2.17	−0.08	0.416
Educator-reported PSC-17						−0.53	0.18	−0.30	0.004*
Block 3:	0.17	0.06*	4.58	4	0.002**				
Age						−0.19	0.13	−0.15	0.146
Gender						-2.41	2.12	−0.12	0.258
Educator-reported PSC-17						−0.36	0.19	−0.20	0.060
Parent-reported PSC-17						−0.67	0.26	−0.27	0.013*
CBCL									
Block 1:	0.01	0.01	0.61	2	0.546				
Age						0.10	0.28	0.04	0.730
Gender						4.91	4.52	0.11	0.279
Block 2:	0.41	0.39**	20.47	3	< 0.001**				
Age						−0.12	0.22	−0.05	0.584
Gender						9.66	3.58	0.23	0.008*
Parent-reported PSC-17						3.20	0.42	0.64	< 0.001**
Block 3:	0.41	0.00	15.19	4	< 0.001**				
Age						−0.13	0.22	−0.05	0.569
Gender						9.58	3.62	0.22	0.010*
Parent-reported PSC-17						3.24	0.45	0.65	< 0.001**
Educator-reported PSC-17						−0.06	0.32	−0.02	0.846
CBCL									
Block 1:	0.01	0.01	0.61	2	0.546				
Age						0.10	0.28	0.04	0.730
Gender						4.91	4.52	0.11	0.279
Block 2:	0.06	0.05*	1.88	3	0.139				
Age						0.16	0.27	0.06	0.562
Gender						6.47	4.50	0.15	0.154
Educator-reported PSC-17						0.78	0.37	0.22	0.039*
Block 3:	0.41	0.35**	15.19	4	< 0.001**				
Age						−0.13	0.22	−0.05	0.569
Gender						9.58	3.62	0.22	0.010*
Educator-reported PSC-17						−0.06	0.32	−0.02	0.846
Parent-reported PSC-17						3.24	0.45	0.65	< 0.001**

Tables presents a summary of hierarchical regression models for baseline educator-report PSC-17 and baseline parent-report PSC-17 total scores as predictor variables with CGAS and CBCL total scores as outcome variables ($n = 92$)

*Significant at $p < 0.05$, ** $p < 0.01$

41]. Findings from the present study that both PSC measures were acceptable to parents and educators support the use of these measures with caregivers of preschool-age children.

Study strengths include the large, national sample complemented by a paired parent and educator sample. This was the first to validate an educator-report version of the PPSC and utilise best-available, gold-standard criterion such as diagnostic interview. However, the large sample could not be used for all analyses and sub-samples were small. Future research may include random sampling to ensure the generalisability of results, and inviting all educators to complete follow-up questionnaires. Community recruitment meant few clinical cases were identified in the first study, which affected predictive validity analysis. Further, parent-reported CBCL may have been a source of criterion contamination; results may have biased parent-rated predictors. This study did not include either a comprehensive educator-report questionnaire nor a clinical assessment including interviews with educators, which may have biased results towards parent reports and not sufficiently accounted for MH symptoms displayed outside the home context. Future research may mitigate this by employing comprehensive educator-report measures or having clinicians interview with educators to inform diagnostic ratings. Future studies may seek to recruit larger educator samples and replicate this study's findings.

In terms of research implications, our findings support the PPSC as a valid, reliable and acceptable measure of child MH for preschool-age children and add weight to the suitability of using the PSC-17 in the preschool-age range with educator report. As free, brief measures, the PPSC and PSC-17 are important offerings in longitudinal research, with clear benefits enabling children to be tracked longitudinally using a single measure.

The first study found that a sizable proportion of children were at risk using the existing PPSC clinical threshold of 9 [10] suggesting this threshold is too low and may overestimate children at risk of MH problems. The existing clinical threshold identified 47.4% of preschool-age children as at risk for MH problems whereas the most recent National Survey of Mental Health and Wellbeing in Australia identified 14% of children aged 4–17 years as scoring in the clinical range of the CBCL [56]. A similar survey in Western Australia which compared prevalence in older children (12–16 years) to younger children (4–11 years) reported 16% prevalence in younger children [57]. The alternative PPSC threshold of 19 identified 10.0% of the sample as at risk. Thus, this revised clinical threshold appears to be more appropriate to use in young Australian children, and should be adopted in future research and clinical practice using the PPSC. Another important implication for determining appropriate thresholds is the level of available healthcare resources that can manage the number of children identified by screening measures as requiring early intervention [58]. Without sufficient resources to manage identified at-risk children, iatrogenic effects such as family burden and over-stretched healthcare systems, may arise as a result of widespread screening programs [58].

Current universal screening and assessment practices in the preschool-age range focus predominantly on school readiness and academic competency [59, 60], however, there is a clear need for screening to be extended to child MH given the high rates of children at risk of problems. The clinical relevance and potential application of the overall findings for both PSC measures indicate caregivers may be willing to adopt screening measures in their educational and care practice. ECEC services, which are less encumbered by issues of access and stigma [61], may be an ideal setting for early identification and can make use of educators as crucial early identifiers of child MH problems. Providing educators with validated tools as part of wider early intervention programs may help identify children in need of support. However, additional training programs are also needed to help educators communicate with parents about child wellbeing and supporting further help seeking [62]. Training educators to support children's MH and to support appropriate parental help-seeking for children is essential given parents frequently seek help for their child's MH problems from a preschool staff member in the first instance [27]. Information about the child from the home and education setting not only contributes to an ideal model of assessment but also may help parents recognise the need for further support for their child's wellbeing, although further research is required.

This research has expanded the evidence base for the validity and acceptability of parent- and educator-reported PPSC and PSC-17 as utilised with young children. It presents new evidence using multi-informant reports, justifying the collection of data from educators. By validating these measures and establishing normative data, this research has increased the clinical utility of the PSC measures, enabling the early identification of young children at risk of MH problems.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13034-025-00985-3.

Supplementary material 1.

Acknowledgements
Not applicable.

Author contributions
RM designed the study, analysed and interpreted the data and drafted the manuscript. LT was involved in the conception and the design of the study, reviewed the data, and reviewed the manuscript. MD was involved in the conception and the design of the study, reviewed the data, and reviewed the manuscript. All authors contributed to the development of the research questions; read, reviewed and approved the final manuscript.

Funding
This work was supported by Growing Minds Australia: A National Trials Strategy to Transform Child and Youth Mental Health Services (Medical Research Future Fund: MRF2006438).

Data availability
The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The University of Sydney Ethics Committee approved both studies (2023/704). Participants provided informed consent prior to participation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 26 June 2025 / Accepted: 14 October 2025
Published online: 09 December 2025

References

1. Mulraney M, Coghill D, Bishop C, Mehmed Y, Sciberras E, Sawyer M, et al. A systematic review of the persistence of childhood mental health problems into adulthood. Neurosci Biobehavioral Reviews. 2021;129:182–205.
2. Lawrence D, Johnson S, Hafekost J, Boterhoven de Haan K, Sawyer M, Ainley J, et al. The mental health of children and adolescents: report on the second Australian child and adolescent survey of mental health and wellbeing. Canberra, ACT: Department of Health; 2015.
3. Gudmundsson OO, Magnusson P, Saemundsen E, Lauth B, Baldursson G, Skarphedinsson G, et al. Psychiatric disorders in an urban sample of preschool children. Child Adolesc Mental Health. 2013;18(4):210–7.
4. Egger HL, Angold A. Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. J Child Psychol Psychiatry. 2006;47(3–4):313–37.
5. Oh E, Mathers M, Hiscock H, Wake M, Bayer J. Professional help seeking for young children with mental health problems. Australian J Psychol. 2015;67(3):187–95.
6. Caspi A, Moffitt TE, Newman DL, Silva PA. Behavioral observations at age 3 years predict adult psychiatric disorders: longitudinal evidence from a birth cohort. Arch Gen Psychiatry. 1996;53(11):1033–9.
7. Jones DE, Greenberg M, Crowley M. Early social-emotional functioning and public health: the relationship between kindergarten social competence and future wellness. Am J Public Health. 2015;105(11):2283–90.
8. Klein AM, Otto Y, Fuchs S, Reibiger I, von Klitzing K. A prospective study of behavioral and emotional symptoms in preschoolers. Eur Child Adolesc Psychiatry. 2015;24(3):291–9.
9. Lipkin PH, Macias MM. Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. Pediatrics. 2020;145(1).
10. Sheldrick R, Henson BS, Merchant S, Neger EN, Murphy JM, Perrin EC. The preschool pediatric symptom checklist (PPSC): development and initial validation of a new social/emotional screening instrument. Acad Pediatr. 2012;12(5):456–67.
11. Gardner W, Murphy M, Childs G, Kelleher K, Pagano M, Jellinek M. The PSC-17: A brief pediatric symptom checklist with psychosocial problem subscales. Ambul Child Health. 1999;5(3):225–36.
12. Spitzer RL, Kroenke K, Williams JB, the Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. JAMA. 1999;282(18):1737–44.
13. Spence SH, Rapee R. Spence Children's Anxiety Scale Brisbane. 1999. https://www.scaswebsite.com/1_1_.html
14. Albers CA, Glover TA, Kratochwill TR. Where are we, and where do we go now? Universal screening for enhanced educational and mental health outcomes. J Sch Psychol. 2007;45(2):257–63.
15. Humphrey N, Wigelsworth M. Making the case for universal school-based mental health screening. Emotional Behav Difficulties. 2016;21(1):22–42.
16. Burkett K, Brown CM, Pickler R, Stanton-Chapman T, Sharps P, Jacquez F, et al. Missed opportunities for intervening early in preschoolers with developmental concern: perspectives from head start parents, teachers, and healthcare providers. J Dev Behav Pediatr. 2025;46(1):e45–55.
17. Honda H, Sasayama D, Niimi T, Shimizu A, Toibana Y, Kuge R, et al. Awareness of children's developmental problems and sharing of concerns with parents by preschool teachers and childcare workers: the Japanese context. Child: Care, Health & Development; 2023.
18. De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DAG, Burgers DE, et al. The validity of the multi-informant approach to assessing child and adolescent mental health. Psychol Bull. 2015;141(4):858–900.
19. Barbarin O, Hitti A, Brown J. Assessing the severity of concerns about preschool children's self-regulation of attention, behavior, and emotions using the ABLE universal screener: A Rasch analysis. J Emotional Behav Disorders. 2020;28(3):167–79.
20. Becker-Haimes EM, Tabachnick AR, Last BS, Stewart RE, Hasan-Granier A, Beidas RS. Evidence base update for brief, free, and accessible youth mental health measures. J Clin Child Adolesc Psychol. 2020;49(1):1–17.
21. McLean RK, Tully LA, Brinley SK, Carl T, Turnell A, Northam JC, et al. Multi-informant universal mental health screening for preschool-aged children by parents and educators: A PRISMA systematic review. Clin Child Fam Psychol Rev. 2024;28:1–21.
22. Stormont M, Stebbins MS. Preschool teachers' knowledge, opinions, and educational experiences with attention Deficit/Hyperactivity disorder. Teacher Educ Special Educ. 2005;28(1):52–61.
23. Franck K, Seland M, Rimul J, Sivertsen AH, Kernan M. Assessing children's psychosocial well-being: Norwegian early childhood education and care teachers' challenges when completing a global screening tool. Contemp Issues Early Child. 2024;25(4):445–59.
24. Brener S, Wright A. Paediatricians' perspectives surrounding identification and management of early childhood disorders in the united States. Child Care Health Dev. 2025;51(1):e70041.
25. Reinke WM, Stormont M, Herman KC, Puri R, Goel N. Supporting children's mental health in schools: teacher perceptions of needs, roles, and barriers. School Psychol Q. 2011;26(1):1–13.
26. Steed E, Banerjee R. Assessment and early identification of young children with social emotional difficulties and behavioral challenges. J Intellect Disabil - Diagnosis Treat. 2016;3:198–204.
27. Pavuluri MN, Luk S-L, McGee ROB. Help-seeking for behavior problems by parents of preschool children: A community study. J Am Acad Child Adolesc Psychiatry. 1996;35(2):215–22.
28. De Los Reyes A, Wang M, Lerner MD, Makol BA, Fitzpatrick OM, Weisz JR. The operations triad model and youth mental health assessments: catalyzing a paradigm shift in measurement validation. J Clin Child Adolesc Psychol. 2023;52(1):19–54.
29. Makol BA, Youngstrom EA, Racz SJ, Qasmieh N, Glenn LE, De Los Reyes A. Integrating multiple informants' reports: how conceptual and measurement models May address long-standing problems in clinical decision-making. Clin Psychol Sci. 2020;8(6):953–70.
30. Lavigne JV, Meyers KM, Feldman M. Systematic review: classification accuracy of behavioral screening measures for use in integrated primary care settings. J Pediatr Psychol. 2016;41(10):1091–109.
31. McLean RK, Tully LA, Dadds M. Reliability, predictive validity and normative data for the pediatric symptom Checklist-17 (PSC-17) in a National Australian sample. Australian New Z J Psychiatry. 2025;59(8):702–12.
32. Liu J, Burgess Y, DiStefano C, Pan F, Jiang N. Validating the pediatric symptoms Checklist–17 in the preschool environment. J Psychoeducational Assess. 2020;38(4):460–74.
33. Gao R, Raygoza A, Distefano C, Greer F, Dowdy E. Assessing measurement equivalence of PSC-17 across teacher and parent respondents. School Psychol Int. 2022;43(5):477–95.

34. Moore SA, Dowdy E, Fleury I, DiStefano C, Greer FW. Comparing informants for mental health screening at the preschool level. School Psychol Rev. 2022;51(5):589–608.
35. DiStefano C, Liu J, Burgess Y. Investigating the structure of the pediatric symptoms checklist in the preschool setting. J Psychoeducational Assess. 2017;35(5):494–505.
36. Leiva L, Antivilo-Bruna A, Peña F, Scquicciarini AM, Simonsohn A, Perrin E, et al. Validation of a measurement model to identify socio-emotional difficulties in preschool children: the preschool pediatric symptom Checklist—Chilean version. Child Psychiatry & Human Development; 2024.
37. Suman P, Mushtaq I, Tandon R. Validation of Hindi translation of survey of Well-Being of young children tool in Indian children. Indian Pediatr. 2022;59(7):539–42.
38. Rocha B, Nunes C. Psychometric characteristics of the Portuguese version of the preschool pediatric symptom checklist for children aged 18 to 60 months. Psicologia. 2021;35(2):109–18.
39. Sheldrick RC, Birudavol J, Hirshfeld-Becker DR, Carter AS, Perrin EC, Dutta A, et al. Longitudinal stability and predictive value of socio/emotional/behavioral screening in an outpatient pediatric network. Acad Pediatr. 2025;25(3):102618.
40. Sheldrick RC, Benneyan JC, Kiss IG, Briggs-Gowan MJ, Copeland W, Carter AS. Thresholds and accuracy in screening tools for early detection of psychopathology. J Child Psychol Psychiatry. 2015;56(9):936–48.
41. Brinley S, Tully LA, Carl T, McLean RK, Cowan CSM, Hawes DJ, et al. Universal child mental health screening for parents: A systematic review of the evidence. Prev Sci. 2024;25:798–812.
42. Harrison JR, Vannest KJ, Reynolds CR. Social acceptability of five screening instruments for social, emotional, and behavioral challenges. Behav Disorders. 2013;38(3):171–89.
43. IBM SPSS. Statistical Package for the Social Sciences. 29.0.1.0 ed. Chicago, IL2023.
44. Australian Bureau of Statistics. Regional population. Canberra; 2022-23.
45. Australian Bureau of Statistics. Household and families: Census. 2021. https://www.abs.gov.au/statistics/people/people-and-communities/household-and families-census/latest-release
46. Australian Bureau of Statistics. New Census insights on income in Australia using administrative data. 2023. https://www.abs.gov.au/articles/new-census-insightsincome-australia-using-administrative-data
47. Hawes DJ, Lechowicz M, Roach A, Fisher C, Dadds MR. Capturing the developmental timing of adverse childhood experiences: the adverse life experiences scale. Am Psychol. 2021;76(2):253–67.
48. Achenbach TM. The Child Behavior Checklist and related instruments. In: Maruish ME, editor. The use of psychological testing for treatment planning and outcomes assessment, 2nd ed. s. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 1999. pp. 429 – 66.
49. Youngstrom EA, Van Meter A, Frazier TW, Hunsley J, Prinstein MJ, Ong ML, et al. Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. Clin Psychol Sci Pract. 2017;24(4):331–63.
50. De Los Reyes A, Langer DA. Assessment and the journal of clinical child and adolescent psychology's evidence base updates series: evaluating the tools for gathering evidence. J Clin Child Adolesc Psychol. 2018;47(3):357–65.
51. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods. 2009;41(4):1149–60.
52. Tissue AD, Hawes DJ, Lechowicz ME, Dadds MR. Reliability and validity of the DSM-5 diagnostic interview schedule for Children, Adolescents, and Parents-5 in externalizing disorders and common comorbidities. Clin Child Psychol Psychiatry. 2022;27(3):870–81.
53. Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, et al. A children's global assessment scale (CGAS). Arch Gen Psychiatry. 1983;40(11):1228–31.
54. Aylward GP. Conceptual issues in developmental screening and assessment. J Dev Behav Pediatr. 1997;18(5):340–9.
55. Reardon T, Spence SH, Hesse J, Shakir A, Creswell C. Identifying children with anxiety disorders using brief versions of the spence children's anxiety scale for children, parents, and teachers. Psychol Assess. 2018;30(10):1342–55.
56. Sawyer MG, Arney FM, Baghurst PA, Clark JJ, Graetz BW, Kosky RJ, et al. The mental health of young people in australia: key findings from the child and adolescent component of the National survey of mental health and Well-Being. Australian New Z J Psychiatry. 2001;35(6):806–14.
57. Garton AF, Zubrick SR, Silburn SR. The Western Australian child health survey: A review of what was found and what was learned. Australian Educational Dev Psychol. 1998;15(1):34–44.
58. Sheldrick RC, Garfinkel D. Is a positive developmental-behavioral screening score sufficient to justify referral? A review of evidence and theory. Acad Pediatr. 2017;17(5):464–70.
59. Cook CR, Volpe RJ, Livanis A. Constructing a roadmap for future universal screening research beyond academics. Assess Effective Intervention. 2010;35(4):197–205.
60. Sim F, Thompson L, Marryat L, Ramparsad N, Wilson P. Predictive validity of preschool screening tools for Language and behavioural difficulties: A PRISMA systematic review. PLoS ONE. 2019;14:e0211409.
61. Desta M, Deyessa N, Fish I, Maxwell B, Zerihun T, Levine S et al. Empowering preschool teachers to identify mental health problems: A task-sharing intervention in Ethiopia. Mind, Brain and Education. 2017;11(1):32–42.
62. Paton K, Gillam L, Warren H, Mulraney M, Coghill D, Efron D, et al. How can the education sector support children's mental health? Views of Australian healthcare clinicians. PLoS ONE. 2022;17(1):e0261827.

Publisher's note