Australian National University

TECHNICAL REPORT:

MEASURING CHILD MENTAL HEALTH, PSYCHOLOGICAL DISTRESS, AND SOCIAL AND EMOTIONAL WELLBEING IN THE LONGITUDINAL STUDY OF INDIGENOUS CHILDREN

Katherine A. Thurber, Jennie Walker, Terry Dunbar, Jill Guthrie, Alison Calear, Philip Batterham, Alice Richardson, Lyndall Strazdins, Maggie Walter, Kate Doery, Raymond Lovett

Standpoint

This work is a collaboration between Aboriginal and non-Indigenous researchers, and is conducted in partnership with Aboriginal and Torres Strait Islander and non-Indigenous research partners. This work is led through the Aboriginal and Torres Strait Islander Health Program at the National Centre for Epidemiology and Population Health, Research School of Population health, Australian National University. Our approach across the Program is to conduct research in partnership with Aboriginal and Torres Strait Islander individuals, communities, and organisations, and to frame our work using a strengths-based approach, where possible. Accordingly, in this Report, we apply decolonising methodologies, including use of strengths-based approaches.

Table of Contents

1. Introduction

The Longitudinal Study of Indigenous Children (LSIC)

Footprints in Time, the Longitudinal Study of Indigenous Children (LSIC) is a national cohort of Aboriginal and Torres Strait Islander children and their parents or carers. Details of the survey are provided elsewhere.[1] In brief, the study was established to answer the primary research question of, 'What do Aboriginal and Torres Strait Islander (Indigenous) children need to have the best start in life to grow up strong?' Two cohorts of children, a B Cohort aged 0-2 years and a K Cohort aged 3-5 years, were recruited at study baseline in 2008. Surveys are conducted annually face-to-face between an Aboriginal and/or Torres Strait Islander interviewer and the Study Child (SC), a parent or carer (Parent 1 or P1), a second parent or carer (Parent 2 or P2) and the child's teacher; teachers were interviewed at some waves. Parent 1 is the parent or carer interviewed about the study child, and as such their responses are included in the current report. The P1 may be the study child's birth mother, a non-birth parent, another family member, or a foster carer. The study includes a total of 1,759 children, and has maintained a high retention rate over the first ten waves of the survey. LSIC is managed by the Australian Government Department of Social Services, and is governed by an Indigenous-majority Steering Committee, who play a key role in the study's design and implementation.[2]

Mental health and social and emotional wellbeing among Aboriginal and Torres Strait Islander children

Mental health is more than the absence of psychopathology; as defined by the World Health Organisation, mental health is 'a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community'.[3]

Many Indigenous peoples internationally hold holistic views of health; for example, Aboriginal and Torres Strait Islander health has been defined as 'not just the physical wellbeing of the individual, but the social, emotional and cultural wellbeing of the whole community … [and] a matter of determining all aspects of their life, including control over their physical environment, of dignity, of community self-esteem and of justice. It is not merely a matter of the provision of doctors, hospitals, medicines or the absence of disease and incapacity'.[4 p. x] Accordingly, many Aboriginal and Torres Strait Islander peoples view the broader construct of social and emotional wellbeing (SEWB) as more relevant than narrower notions (i.e. mental health and the absence of pathology). SEWB is a strengths-based concept inclusive of positive mental health; SEWB is tied to other domains of wellbeing including physical health, spirituality, and culture, and also to connection to family, community, and land.[5,6] This Report will consider the extent to which the mental health-related measures examined in this report may be relevant as indicators of SEWB more broadly.

It is well-established that Aboriginal and Torres Strait Islander children experience a disproportionate burden of poor mental health and associated harms,[7] stemming from the history of colonisation and dispossession, and ongoing trauma.[8-10] Despite the great need, there is a paucity of measures of mental health and/or SEWB that are evidenced to be culturally appropriate and valid for use with Aboriginal and Torres Strait Islander peoples.[11-14] In particular, there is an absence of evidence on the validity and appropriateness of measures in different contexts.[13]

While numerous measures of mental health have been validated for use with non-Indigenous Australian children, it cannot be assumed that these measures are valid for use with Aboriginal and Torres Strait Islander children, i.e. that they have cross-cultural validity.[14-17] It is well-established that

there is cultural variation in the experience and expression of mental health conditions and in the types of behaviours considered 'problematic', including between Indigenous and non-Indigenous populations.[14,18-21] For example, Dingwall and Cairney write, 'Definitions of what constitutes normal (or abnormal) expressions of behaviour can vary considerably between cultures and even within cultures … Psychiatric disorders may also be expressed through different symptomatology. For example, anger may represent a culturally specific symptom of depression for Indigenous Australians'.[11 p. 467] In the context of First Nations peoples in the United States, Pace et al. write, 'there is evidence to suggest that some behaviors that are viewed as signs of distress or maladjustment in mainstream contexts are often viewed as evidence of successful adaptation to unique stressful circumstances. In other cases, these types of behaviors may be seen as indications of spiritual advancement and wisdom'.[21 p. 331] As a result of these differences, Dauphinais et al. argue, 'Until further test validation research is accomplished, most tests, when used cross-culturally, should be assumed to be biased'.[15 p. 107]

While it is well-established that there is cultural variation in the experience and expression of mental health conditions (including between Indigenous and non-Indigenous populations), we currently lack a detailed understanding of the lived experience of mental health and disorders among Aboriginal and Torres Strait Islander peoples; 'There remains a need for Indigenous voices to be heard in order to explore and gain greater knowledge of their conceptualizations of mental disorders and to consider these in relation to Western biomedical conceptualizations'.[9 p. 476] Further, it is unknown how understandings of mental health vary by context (i.e. by mob, level of remoteness, or other). Given the vast heterogeneity of the Aboriginal and Torres Strait Islander population, it is likely that conceptualisations of mental health vary between groups.[9,13,14,22]

Assessment of mental health and/or SEWB using measures developed for other populations can lead to misdiagnosis and cultural bias if they are measuring constructs or symptoms that are not relevant to the population of interest.[9,13,14,21,23] The importance of culturally-specific assessment of mental health and SEWB was acknowledged in 1995 in the landmark report, *Ways Forward*, the first national assessment of Aboriginal and Torres Strait Islander mental health and SEWB.[10] The report outlined guiding principles for the development of a Mental Health Strategy and Plan, including: 'Culturally valid understandings must shape the provision of services and must guide assessment, care and management of Aboriginal and Torres Strait Islander peoples' health problems generally, and mental health problems in particular'.[10 p. 19] This principle remains in the current National Strategic Framework for Aboriginal and Torres Strait Islander Peoples' Mental Health and Social and Emotional Wellbeing (2017-2023).[24] In a 2017 systematic review of SEWB measures for Aboriginal and Torres Strait Islander peoples, Le Grande et al. conclude: 'It is advised that standard instruments only be used if they have been subject to a formal cross-cultural adaptation process, and Indigenous developed measures continue to be developed, refined, and validated within a diverse range of research and clinical settings'.[14 p. 164]

Identification and/or development of robust measures of mental health and SEWB for Aboriginal and Torres Strait Islander peoples is urgently required, to support improvements to mental health and SEWB and related services.[11,14] This has been identified as a policy priority; for example, 'Develop culturally appropriate indicators to measure social and emotional wellbeing' is identified as a key strategy in the National Strategic Framework for Aboriginal and Torres Strait Islander Peoples' Mental Health and Social and Emotional Wellbeing (2017-2023).[24 p. 18]

Assessment of the standard psychometric properties of a measure is not sufficient to demonstrate that a measure is valid for use with Aboriginal and Torres Strait Islander peoples: 'Levels of Indigenous cultural validation, community engagement, and compliance with Indigenous knowledge-making

processes that occur when developing Indigenous assessment instruments should also be considered'.[13 p. 47] Newton et al. argue that these factors need to be incorporated into the definition of a 'gold standard' measure.[13] Accordingly, in this Report, we take these factors into consideration, alongside assessment of psychometric properties.

The purpose of this report

Dingwall and Cairney explain the challenge of identifying measures fit for use with the Aboriginal and Torres Strait Islander population: 'An appropriate assessment must therefore meet two often conflicting prerequisites: it must be both scientifically and biomedically valid; and it must have relevance culturally for the target group'.[11 p. 22]

The purpose of this Report is to assess the validity of measures relevant to mental health, psychological distress, and/or social and emotional wellbeing of Aboriginal and Torres Strait Islander children in LSIC. In doing so, this Report contributes to expanding the evidence base surrounding the validity of measures of mental health, psychological distress, and SEWB for Aboriginal and Torres Strait Islander children. This Report assesses the appropriateness of measures developed for other populations for use with Aboriginal and Torres Strait Islander children, and also assesses the appropriateness of a measure developed with and for Aboriginal and Torres Strait Islander youth.

The structure of this report

Chapter 2 provides an overview of the measures that will be explored in this Report. Chapter 3 describes the overarching methodology and analytical methods used in the Report. An assessment of each measure is provided in Chapters 4-9, along with recommendations for data users.

2. Overview of child mental health-related measures used in LSIC

Table 2.1 provides an overview of the measures relevant to mental health in LSIC. This includes the Strengths and Difficulties Questionnaire (SDQ), Strong Souls: Strengths, Social Skills Improvement System (SSIS), Short Infant Temperament Questionnaire (SITQ), School-Age Temperament Inventory (SATI), and Brief Infant Toddler Social Emotional Assessment (BITSEA). These are intended to measure a range of constructs relevant to mental health, including social and emotional behavioural difficulties, temperament, social skills, and resilience. With the exception of the SDQ, LSIC includes subsets of each measure. The measures are designed for use with children of ages varying from 1-18 years; the measures are administered to the relevant age groups, as outlined in Table 2.1 and Table 2.2.

Table 2.1. Overview of child mental health-related measures in LSIC

Measure	Intended construct	Full or partial measure	Age group (years)
Strengths and Difficulties Questionnaire (SDQ)	Social and emotional behavioural difficulties.	Full	2-17
Strong Souls: Strengths	Resilience	Partial	Not stated
Social Skills Improvement System (SSIS)	Social skills and problem behaviours	Partial	3-18
Short Infant Temperament Questionnaire (SITQ)	Temperament and social skills: cooperation, assertion, responsibility, empathy, self-control, communication, engagement	Partial	3-8
School-Age Temperament Inventory (SATI)	Temperament / personality: negative reactivity, task persistence, approach, activity	Partial	8-11
Brief Infant Toddler Social Emotional Assessment (BITSEA)	Social-emotional and behavioural problems	Partial	1-3

Table 2.2 provides details on the mental health-related measures used in LSIC, according to informant (P1-report, teacher-report, child self-report), number of items, variable names in the dataset, and the waves and cohorts (and corresponding ages) at which the measures are administered. The SDQ is administered most frequently; the other measures are administered at one or two time points. The SDQ is administered to all three informants at different time points; SSIS is the only other measure examined that includes child self-reported items.

Detail on each measure is provided in subsequent chapters. For each measure, the following is described: the purpose of and rationale for the measure; any previous evidence on the validity of the measure for the Aboriginal and/or Torres Strait Islander population; use of the measure with Aboriginal and/or Torres Strait Islander population; questionnaire items and response options; coding and scoring; and, modifications between the original measure and the measure used in LSIC.

Table 2.2. Child mental health-related measures in LSIC

Measure	# items	Variable names	Age of B Cohort									
			0-2	1-3	2-4	3-5	4-6	5-7	6-8	7-9	8-10	9-11
		Age of K Cohort	3-5	4-6	5-7	6-8	7-9	8-10	9-11	10-12	11-13	12-14
			W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
SDQ: parent- or carer-reported	25	asqemot asqcond asqhype asqpeer asqdiff asqpros			✓	(K)		✓		✓		✓
SDQ: child self-report	25	csqemot csqcond csqhype csqpeer csqpros csqdiff										(K)
SDQ: teacher-reported	25	dsqemot dsqcond dsqhype dsqpeer dsqdiff dsqpros		(K)	✓	✓	✓	✓			✓	✓
Strong Souls: Strengths	12	css1_a to css1_m									(K)	
Social Skills Improvement System (SSIS): parent- or carer-reported	21	asr1_a to asr1_f, asr2_a to asr2_d, asr3_a to asr3_c, asr3_e, asr4_a to asr4_e										✓
Short Infant Temperament Questionnaire (SITQ)	12	apa4_a to apa4_l		(K)			(B)					
Social Skills Improvement System (SSIS): child-reported	10	csr5_1 to csr5_6										(K)
School-Age Temperament Inventory (SATI)	12	apa8_a to apa8_l						(K)		(B)		
Brief Infant Toddler Social Emotional Assessment (BITSEA)	34	apa51 to apa 534		(B)								

SC = Study Child. Approximate age ranges of Study Children at each Wave are presented in years. Questions are asked in relation to all children, unless indicated as B Cohort (B), K Cohort (K), or new parents or carers only (N). The prefix corresponding to study wave is removed from variable names.

3. Methods

Research methodology

> *Research constructs are enmeshed in the culture and worldview of Euro-American society and serve to support its continuation by judging which research constructs are valid, determining how constructs are defined, and deciding which variables need to be controlled.*[25 p. 123]

This project employs decolonising methodologies: it adopts a participatory and strengths-based approach; it is guided by Indigenous worldviews and ways of knowing; it is grounded in respect for cultural history, knowledge, and protocols; it recognises and values the knowledge of community members; and it is committed to Indigenous self-leadership of research.[25-27] Decolonising approaches are important given the history of exploitive research practices with Aboriginal and Torres Strait Islander communities, and Indigenous populations internationally.[17,22,28-32] Use of these methodologies, rather than standard methodologies, is 'more likely to provide congruency between Aboriginal and Torres Strait Islander peoples' understanding of [mental health and SEWB] and the development of instruments designed to measure it'.[22 p. 192]

Participatory research is designed to enable accurate portrayal of Aboriginal and Torres Strait Islander peoples and communities, and to generate findings that are useful for participants and their communities. Participatory research is founded on collaboration between researchers and community, recognising and valuing the knowledge of community members. Participatory research entails co-development of research aspects such as research questions, study design, data collection, interpretation, dissemination, and translation.[2]

This project is led and governed by Aboriginal and Torres Strait Islander researchers, and is conducted in partnership with Aboriginal and Torres Strait Islander stakeholders through ongoing consultation processes. Briefly, the analysis was co-designed and interpreted by Aboriginal and Torres Strait Islander researchers and non-Indigenous researchers. A knowledge exchange session was held with eight LSIC Research Administration Officers (RAOs) in February 2019. The key aims of this session were: to learn from RAOs' experience administering these measures in the field, to guide analysis, and to ensure that findings and interpretations were aligned with their views.[27] Preliminary findings were presented to the LSIC Steering Committee in April 2019, with additional input sought via email. Learnings from these exchanges are incorporated into the following chapters.

General analytical approach

For each of the measures of interest, we applied a general analytical approach, as described below. The exact text of each item within each measure is presented in this Report, except where copyright agreements prohibit replication of the exact items. In this case the name of the variable in LSIC corresponding to this item is presented; the first letter of the variable name, which identifies the wave of the study in which the variable is used, is omitted from the variable name for ease of interpretation.

First, we reviewed and synthesised the existing evidence on the measures and their validity for the Aboriginal and Torres Strait Islander population, or for other populations (e.g., all Australian children, or Indigenous populations internationally) where evidence specific to the Aboriginal and Torres Strait Islander population is not available. This identified the specific knowledge gaps to be filled through analysis of data from LSIC.

Acceptability: We quantified the extent of non-response across items and the total scale score, as an indication of the acceptability of the items and scale. A greater extent of non-response was interpreted

as evidence of potential item unsuitability. Non-response across items and measures was summed and presented as a rate per 100 surveys. Where the extent of non-response was low, both cohorts were combined and a single non-response rate was reported, as relevant.

Construct validity: Exploratory and/or confirmatory approaches were conducted to assess if the scale was measuring the intended underlying construct(s) or factor(s). Exploratory methods (Principal Components Analysis, PCA) is used where there was not an *a priori* hypothesised structure or number of dimensions underlying the measure; Confirmatory Factor Analysis (CFA) is used when testing a specific hypothesis. We interpreted the loading of items on each component, and where possible labelled the component. Components were presented in the order of the amount of variance they explained in the sample; for example, in all cases, Component 1 explained the highest amount of variance in the sample of all the components. Results were compared to the hypothesised structure of the measure, and to previous research from the reference population and/or specific to the Aboriginal and/or Torres Strait Islander population. These processes assisted in making recommendations to reduce the number of items in a measure.

Before undertaking PCA, suitability of the data for PCA was assessed. The relationship between the items, by informant, was tested using a Pearson correlation coefficient (r). Pearson correlation ranges from -1 to 1 with greater scores (in absolute value) indicating a stronger relationship between the variables. Items with a low (absolute value of r, $|r| < 0.3$) or high ($|r| > 0.8$) correlation may be unsuitable for a PCA, and consequently may be excluded.[33] Throughout the report, where we refer to the size of a correlation (r), we refer to the magnitude of the correlation in absolute value, i.e. ($|r|$). A significant score on the Bartlett's test of sphericity and a Kaiser-Meyer-Olkin (KMO) score above 0.6 was used to assess if item correlations were large enough to be suitable for PCA.[34]

For PCA, the number of components for extraction was determined by a three-criterion approach. First, a parallel analysis was assessed and components with an actual eigenvalue greater than a random order eigenvalue were retained. To generate random order eigenvalues, ten datasets were created at random with the same number of observations and items as the LSIC data in question. The resulting correlation matrices were averaged then the random order eigenvalues were computed. Second, a screeplot was evaluated visually, and the number of components was defined as the number present before the plot shape changed direction. Third, the number of components with an eigenvalue > 1 (Kaiser's criterion) were retained, noting that this approach is understood to generally overestimate the number of components.[33] The final decision was made based on the balance of results from these three criteria.

In general, results from PCA are sample-specific, whereas results from CFA are more generalisable to the total population. Where findings from PCA are consistent with previous evidence from this population, this supports broader generalisability of the findings.

Internal consistency: We undertook an internal consistency assessment of each mental health measure to assess the extent to which items within each measure were related. The internal consistency of SDQ was assessed using Cronbach's alpha. Internal consistency is generally defined as 'acceptable' if the alpha is ≥ 0.7;[35] for the purposes of this report, we considered alpha levels between 0.6 and < 0.7 to be 'approaching acceptable', and alpha levels below 0.6 'unacceptable'. While there is no consistent agreement on what level of alpha is 'too high', it is agreed that very high alphas can indicate item redundancy,[36,37] that is, that the measure is 'asking the same question many different ways'.[38 p. 30] Item redundancy might indicate that some items should be removed; respondents who feel like they are answering multiple questions about the same thing can become frustrated, which

can reduce data quality.[36,37] Streiner recommends a maximum alpha value of 0.90. Accordingly, in this Report, we highlight where the alpha is > 0.90.

Convergent and/or divergent validity: Associations between mental health and potentially related outcomes were quantified to test construct validity (convergent and/or divergent validity). Table 3.1 below outlines potential outcome measures for use in assessing the convergent/divergent validity of child mental health-related measures. Where relevant, correlations between different measures of child mental health were examined. A correlation of *r* = 0.10 to 0.29 was considered 'weak', *r* = 0.30 to 0.49 'moderate' and *r* = 0.50 to 1.00 'strong'.[39]

Table 3.1. Potential outcome measures for validity assessment

Outcome	# items	Variable names	Age of B Cohort 0-2 / Age of K Cohort 3-5 / W1	1-3 / 4-6 / W2	2-4 / 5-7 / W3	3-5 / 6-8 / W4	4-6 / 7-9 / W5	5-7 / 8-10 / W6	6-8 / 9-11 / W7	7-9 / 10-12 / W8	8-10 / 11-13 / W9	9-11 / 12-14 / W10	10-12 / 13-15 / W11*
Anxiety, depression*	1	ahc2_73, ahc2_74											✓
ADD or ADHD	1	ahc2_28	(K)	(K)	✓	✓	✓	✓	✓	(B)	✓		✓
Autism	1	ah2_36		✓	✓	✓	✓	✓	✓	✓	✓		✓
Considered suicide	1	csu3											(K)
Parents' Evaluations of Developmental Status (PEDS)	2	ald10, ald11	✓	✓		(B)	(B)	(B)					

Approximate age ranges of Study Children at each Wave are presented in years. Questions are asked in relation to all children, unless indicated as B Cohort (B), K Cohort (K), or new parents or carers only (N). The prefix corresponding to study wave is removed from variable names.

* Wave 11 beta data on child mental health-related diagnoses and other outcomes are provided for the purpose of this report.

Criterion validity: The performance of the SDQ Total Difficulties and SDQ Prosocial was assessed against a parent- or carer-reported child mental health outcomes 12 months later. Mental health outcome was the reporting of anxiety/depression, ADD/ADHD and/or autism. Ideally the performance of the SDQ would be measured against a 'gold standard', but in the absence of a 'gold standard' the self-reported measure was assessed; in the absence of any 'gold standard' measure, use of these measures is justified.[40] Criterion validity was measured by sensitivity, specificity and the positive predictive value (PPV).

Sensitivity is the percentage of children with true problems (i.e. with any of the included mental health-related conditions) that are correctly categorised by the instrument (i.e. categorised as 'abnormal' by the SDQ); 70-80% is an acceptable range for sensitivity.[41] *Specificity* is the percentage of children without problems (i.e. without any of the included mental health-related conditions) who are correctly categorised by the instrument (i.e. SDQ Total Difficulties Score lower than the cut-off for 'abnormal'); 80% is a desirable level of specificity. *Positive Predictive Value (PPV)* is the percentage of children categorised by the instrument as being 'high risk' that have true problems (i.e. the percentage of those in the 'abnormal' SDQ category that have a mental health condition). While a higher PPV is desirable, there is not an agreed standard for PPV; screening instruments often have PPVs as low as 30-50%.[41]

We calculated *Receiver Operating Characteristics* (ROCs) for the SDQ Total Difficulties score. ROC curves are used to assess the sensitivity and specificity of a measure at each possible cut-off point, and to identify cut-off points that optimise sensitivity and specificity, that is, minimise the total error (false positives and false negatives).[40] The aim of this analysis was to identify cut-off points for the SDQ Total Difficulties Score that optimise sensitivity and specificity in relation to mental health-related conditions in the LSIC sample. Choosing a cut-off is always be a trade-off between sensitivity and specificity;[42] in this Report, we have prioritised sensitivity over specificity when suggesting cut-off

points, as screening tools should aim to minimise false negatives. LSIC data users may choose different cut-off points, depending on the relative importance of sensitivity versus specificity in relation to the research question. We note that the mental health outcomes used are not 'gold standard' measures; we are relying on parent- or carer-report of child mental health outcomes, and child self-report of considering suicide (suicide ideation). However, in the absence of any 'gold standard' measure, use of these measures is justified.

Additional analyses were conducted, as indicated by the results, and as guided by Aboriginal and Torres Strait Islander stakeholders and DSS. This included assessment of differences in psychometric properties across respondents (e.g. by informant, child age, child gender, or level of remoteness).

Based on these findings, recommendations for the use of measures and questionnaire items are provided. Guidance for data users will be provided around the use of items and measures; recommendations will take into account contextual considerations for this population.

Ethics

This Project, including the key informant focus group, was conducted with ethics approval from the ANU Human Research Ethics Committee (HREC, Protocol 2016/534: Social and emotional wellbeing in the Longitudinal Study of Indigenous Children).

4. Strengths and Difficulties Questionnaire (SDQ)

Background

Purpose of the SDQ

The SDQ (Copyright Robert Goodman, 1999, UK) is a 25-item questionnaire designed for use as a screening tool with children aged 2-17 years.[43,44] It asks a range of questions about negative and positive child attributes. It can provide an initial assessment of emotional and behavioural difficulties across four difficulties subscales: Emotional Symptoms, Conduct Problems, Hyperactivity, Peer Problems, and one strengths subscale: Prosocial Behaviour.

Seven versions of the SDQ have been developed to accommodate the age of the child (2-4 years, 4-10 years, 11-17 years) and the type of SDQ informant (parent or carer, teacher, or child); child self-complete SDQ is only available for the 11-17-year age group. Further information on the different SDQ versions and associated scoring can be found at http://www.sdqinfo.com/py/sdqinfo/b0.py. Throughout LSIC the various versions were administered to correspond with the Study Child's age (Table 4.1). However, as B cohort Study Children will be aged 11-13 years at Wave 12 (2019), it is anticipated only the 11-17 years version will be administered in future LSIC waves.

Table 4.1. Version of SDQ administered by wave, cohort, and SDQ informant

Wave	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
B Cohort										
Parent- or carer-report			2-4 years			4-10 years		4-10 years		4-10 years
Teacher-report			2-4 years	2-4 years	4-10 years	4-10 years			4-10 years	4-10 years
Child self-report										
K Cohort										
Parent- or carer-report			4-10 years	4-10 years		4-10 years		4-10 years		11-17 years
Teacher-report		4-10 years	4-10 years	4-10 years	4-10 years	4-10 years			11-17 years	11-17 years
Child self-report										11-17 years

The SDQ is asked of all parents or carers in Waves 3, 6, 8, and 10, and of parents or carers of children in the K Cohort only in Wave 4 (Table 4.1). Additionally, data are available on teacher-reported SDQ (multiple waves) and child self-reported SDQ (Wave 10, K Cohort only).

All SDQ informants were prompted, *For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain*. Parents or carers are asked to *Please give your answers on the basis of the child's behaviour over the last six months or this school year*; teachers are asked to *Please give your answers on the basis of this child / young person's behaviour over the last six months or this school year*; children are asked to *Please give your answers on the basis of how things have been for you over the last six months or this school year.* For each of the 25 items, participants are asked to respond whether the statement is not true (coded as 0), somewhat true (coded as 1), or certainly true (coded as 2).

A score for each subscale is created by summing items responses within the subscale. The subscale is not derived if data are missing for two or more of the items. The SDQ Total Difficulties Score is created by summing scores across the first four subscales; the Prosocial subscale is not included in the Total Difficulties Score, and is intended to be analysed separately. The Total Difficulties Score is only calculated if all four subscales are non-missing.

There are established cut-off points for categorising children's SDQ scores; the cut-offs vary by SDQ age version (2-4 years versus 4-17 years) and by the type of SDQ informant (P1, Teacher, or Child).[45,46] There are multiple categorisation systems currently in use. For example, according to the original three-band cut-points, Total Difficulties Scores for children aged 4-17 years are categorised as 'normal' (0 to 13), 'borderline' (14 to 16), or 'abnormal' (17 to 40) for the P1-completed SDQ; the respective cut-off points are 0-11, 12-15, and 16-40 for the teacher-completed SDQ, and 0-15, 16-19, and 20-40 for the child self-complete SDQ.[45] The different cut-off points for different informants and survey age versions should be considered when comparing SDQ scores from different informants. However, SDQ risk categories are intended to be comparable across versions.

The original three-band cut-points were created based on a UK sample, defined such that 80% of the population would fall in the 'normal' range, and 10% each in the 'moderate' and the 'high' range.[43,47] Alternative categorisations do exist; for example, norms have been calculated for children aged 7-17 years in the total Australian population.[48] To date, the three-band classification system (employing the cut-off points from the normative UK sample) has been most commonly used with the Aboriginal and Torres Strait Islander population. As such, we focus on this classification system in this Report.

While the SDQ is intended to have a five-factor structure, support for this structure in the international literature is mixed, meaning that 'the hypothesised five subscales may not always tap distinct constructs'.[49 p. 1189] The performance of the five-factor structure is particularly questioned in samples where the risk of mental health pathology is low.[49] In this context, Goodman et al. propose the use of broad Internalising subscale (Emotional Symptoms and Peer Problems combined) and Externalising subscale (Conduct Problems and Hyperactivity combined), alongside the Prosocial subscale, instead of using the five standard SDQ subscales.[49] To our knowledge, the five subscale approach is still predominantly used. Research suggests that a lack of fit with the hypothesised five-factor structure of does not undermine the SDQ's utility in predicting clinical disorders.[49,50]

The 'extended' version of the SDQ includes an 'impact supplement', with questions about whether the respondent perceives the child to have any difficulties with emotions, concentrations, behaviour, or relationships.[51] If the respondent answers yes, the respondent is asked how long they think the difficulties have been present, and are asked about the perceived impacts in terms on the child (distress, interfering with home life, friendships, classroom learning, or leisure activities), and the perceived burden on the family. The impact supplement includes a total of one to seven additional questions. A validation study of the impact supplement found that the impact score was a better predictor of clinical status than the standard SDQ symptom scores; the combination of symptom and impact scores was the best predictor of clinical status. In addition, use of the single item *Overall, do you think that your child has difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people?* was a strong predictor of the presence/absence of psychiatric disorder, and only slightly less accurate than use of the combination of symptom and impact scores.[51] The authors conclude, 'Developmental, paediatric, and educational researchers might want to consider including this "one item screen" in their questionnaires when their respondents' time is at such a premium that even a brief screening instrument such as the SDQ would be unacceptably long. Obviously, the "one item screen" provides no hint as to the type of disorder'.[51 p. 797]

In previous qualitative research, Aboriginal and Torres Strait Islander people strongly supported the use of the SDQ impact supplement; 'This section of the SDQ was viewed as clear and straight to the point and participants agreed that the questions contained within it were very important'.[5 p. 897] The SDQ impact supplement questions have also been used with Indigenous populations internationally, including with Māori in New Zealand[52] and Sami in Norway.[20,53]

Items from the SDQ impact supplement were included in questionnaires in the SEARCH study. While two items from the SDQ impact supplement have been examined in relation to parent or carer mental health-related outcomes[54], we are unable to find any published literature examining SDQ scores in relation to the SDQ impact supplement within SEARCH or any other Aboriginal and/or Torres Strait Islander cohort. To our knowledge, the SDQ impact supplement was not used in WAACHS.

The SDQ impact supplement has not been included in any of the first 11 waves of the LSIC survey. One reason that LSIC did not include the impact supplement is that similar questions are asked through the Parents' Evaluations of Developmental Status (PEDS) items administered in LSIC in Waves 1 and 2 (both cohorts) and Waves 3-5 (B Cohort only). PEDS is a brief parent report measure of parent's concern about their child's development, intended to be used as an initial screen to determine a child's risk for various disabilities and delays.[41] Research from other populations has identified that eliciting parental concern about children's development can be as effective as using formal developmental screening tools; 'Parental concerns about language, fine motor, cognitive, and emotional-behavioral development are highly predictive of true problems'.[55 p. 193] The PEDS items have been recommended as questions to elicit parent or carer concern within the clinical setting.[41]

PEDS can be used with children aged 0-9 years; based on data from other populations, it has a sensitivity ranging from 74-79% and specificity from 70-80% across age levels.[41] There are 10 PEDS items in total; LSIC includes the following items from PEDS (P1-reported) which are similar to the SDQ impact supplement questions: *Do you have any concerns (worries) about how SC behaves?* And *Do you have any concerns (worries) about how SC gets along with others*?[1] While these items do not cover the breadth of concerns covered in the impact supplement item (in particular, the PEDS questions do not cover concerns about the child's concentration), there is a large degree of overlap, and it was therefore considered repetitive to additionally include this SDQ impact supplement item. The PEDS items could be used as a proxy measure of the SDQ impact supplement (single item) at waves/cohorts where both the SDQ and PEDS items are asked. At Wave 6, parents or carers of children in the B Cohort completed both the PEDS items and the SDQ.

Evidence on the validity of the measure for Aboriginal and/or Torres Strait Islander children
The Western Australian Aboriginal Child Health Survey (WAACHS) conducted from 2000-2001 employed the SDQ in its survey of the health and wellbeing of 5,289 Aboriginal and Torres Strait Islander children living in Western Australia. Zubrick et al. conducted an assessment of the reliability and internal consistency of the parent-reported SDQ for children aged 4-17 years.[56] WAACHS used a slightly modified version of the SDQ, with permission from Robert Goodman, based on results from their pilot study. The survey was administered face-to-face, rather than paper self-completion given concerns about literacy of participants. The response options were modified based on the pilot testing; not true was changed to no, somewhat true was changed to sometimes, and certainly true was changed to yes. There is no published information about the process of instrument modification.[5]

Overall, they found that the modified SDQ had good internal consistency and validity. The reliability of the subscales was assessed using a method by Raykov, rather than by using Cronbach's alpha. The

---

[1] LSIC received copyright permission to use 'worries' as a clarifying or alternate word, when appropriate. Stakeholders in LSIC considered that 'worries' is a more frequently used and better understood word than 'concerns'. This has been identified by Indigenous stakeholders in previous research.16. Kowal E, Gunthorpe W, Bailie RS. Measuring emotional and social wellbeing in Aboriginal and Torres Strait Islander populations: an analysis of a Negative Life Events Scale. *International journal for equity in health* 2007; **6**(1): 18.

Total Difficulties Score and all subscales except the Peer Problems subscale were deemed robust; the Peer Problems subscale performed more poorly. The scale reliability for the Total Difficulties Score was 0.93. They also identified that while the overall reliability was satisfactory, reliability varied by level of remoteness and was lower in more remote areas. The authors hypothesised that this might be due to factors including the need for translation from English to the local Aboriginal language, or that 'some concepts were less salient to cultural and living circumstances' for remote participants.[56 p. 20] They used CFA to test if the questionnaire items mapped on to the underlying factor (i.e. Emotional Symptoms, Conduct Problems, Hyperactivity, Peer Problems, and Prosocial Behaviour). They found that the majority of the factor loadings were between 0.5-0.8, indicating an 'acceptable' relationship between the items and the factor. They found that the Peer Problems subscale was the least reliable of the five subscales, with factor loadings of < 0.5 for some of the items. They identified some evidence of differences in the interpretation of the SDQ by the child's gender, the age of the child, and characteristics of the responding parent or carer (birth mother compared to non-birth mother; Aboriginal compared to non-Aboriginal). They conclude, 'On balance these findings suggest that the total SDQ score is likely to be an adequate measure of mental health distress'.[56 p. 22]

Despite the extensive validation work undertaken with the modified WAACHS SDQ instrument, the WAACHS modified SDQ has not been used in subsequent research, to our knowledge. The standard version is 'still being mandated for use clinically and employed in most research programmes involving Aboriginal children'.[5 p. 895] The unmodified SDQ is used in LSIC.

Williamson et al. also the examined the acceptability and validity of the SDQ for Aboriginal children, through focus groups with three Aboriginal community controlled health services (ACCHOs), and analysis of quantitative data from the Study of Environment on Aboriginal Resilience and Child Health (SEARCH), a cohort study of children from four ACCHOs areas in urban New South Wales.[5,50] They found that the unmodified SDQ was acceptable for this urban population, but would benefit from some changes to the questions and response scale: 'While it is always advantageous for the purposes of comparison to use the same measurement tool across all groups, this advantage only holds if respondents share a common interpretation of the questions asked and indeed the response scale'.[5 p. 899] The modifications suggested by participants in this study were different to the modifications used in WAACHS. This might reflect geographic variation in the acceptability and interpretability of items.[5] This has potential implications for the acceptability and interpretability of items in the LSIC sample, given the extent of the sample's diversity in terms of geography, remoteness, and other characteristics. Williamson et al. concluded, 'It is likely that modifications to the wording of the SDQ would significantly improve the cultural appropriateness of the SDQ for Aboriginal children in Sydney. It is not necessarily the case, however, that such modifications would improve the diagnostic utility of the tool'.[5 p. 899]

In particular, Williamson et al.[5] identified limitations with the Peer Problems subscale. This subscale is exclusively focused on peer relationships, rather than considering extended family members, and was considered to not 'fit well with Aboriginal concepts of the relative importance of different interpersonal relationships'.[5 p. 894] This finding aligns with the reduced internal consistency of this subscale in WAACHS.[56] Participants also identified that concepts important to mental health were missing from the SDQ. This included: connection to extended family, identity, feeling accepted by and belonging to community, experiences and impacts of racism, respect for Elders, sleeping problems, anger, changes in eating habits, displays of affection, and cruelty to animals.

Williams et al.[50] tested the acceptability, construct validity, internal consistency, and convergent validity of the P1-report SDQ among 717 children aged 4-17 years in the SEARCH cohort. Acceptability was assessed based on the percentage of respondents who completed the SDQ, and who selected

don't know, refused, or missing for each question. They assessed internal consistency of the Total Difficulties Score and each of the subscales using Cronbach's alpha. Internal consistency was acceptable for the Total Difficulties Score (alpha = 0.85), and for all of the subscales except Peer Problems (alpha = 0.47). They conducted CFA to test the hypothesised SDQ factor structure in the whole sample. There were four items with factor loadings that were unacceptably low (< 0.4): *often complains of headaches, stomach aches or sickness*; *rather solitary, tends to play alone*; *picked on or bullied by other children*; and *gets along better with adults than with other children*. Three of these were from the Peer Problems subscale. They retested CFA after removing the Peer Problems subscale, but it did not lead to any material improvements.

They tested the convergent validity of the SDQ Total Difficulties Score by quantifying its relationship with P1-reported problematic behaviours in the past six months: *did not want to go to school; ran away from home; drunk alcohol or gotten drunk; deliberately harmed; talked about death or suicide.* This analysis was restricted to school-aged children. They used the Kruskal-Wallis test for equality of medians. They found that the median SDQ Total Difficulties Score was significantly higher for children engaging in the problematic behaviour compared to not engaging in the behaviour, across all of the outcomes except alcohol use.

Overall, Williamson et al.[50] support use of the SDQ Total Difficulties Score, and recommend not using the Peer Problems subscale. They recommend further validation work for the P1-reported SDQ, including clinical, predictive, discriminant and face validity, and test-retest reliability. This literature also points to the importance of exploring reliability and validity across age groups, sex, and levels of remoteness. Further, to our knowledge, there is no evidence on the reliability or validity of the teacher-reported or child self-reported SDQ for this population.

Use of SDQ with Aboriginal and/or Torres Strait Islander children

The SDQ Total Difficulties Score has been utilised in quantitative analysis of data from WAACHS, SEARCH, and LSIC,[57-62] as well as in other cohorts. Most of these studies have examined factors associated with SDQ score.[56,58,62] Taking a strengths-based approach, many of these papers have focused on the outcome of low rather than high, SDQ Total Difficulties Score, as a measure of 'good mental health' or 'resilience'.[58,61,62] Marmor et al. explored the relationship between measures of SEWB and the SDQ Total Difficulties Score.[57] The authors found that the SDQ Total Difficulties Score was positively associated with measures of connectedness and relationships (key components of SEWB), suggesting SDQ Total Difficulties Score was not an acceptable proxy for SEWB.

Results: assessment of SDQ in LSIC

Acceptability

Non-response

Non-response was assessed for each informant (P1, Teacher, or Child). Within each wave, the extent of non-response was very low. Given this, wave and cohort were combined, acknowledging that the majority of parents or carers and some teachers responded about the same child at multiple waves.

Across the total scale (25 items), teachers had the highest rate of non-response, at 59.8 item non-responses per 100 surveys, followed by children with an average of 51.0 item non-responses per 100 surveys. Across the total scale (25 items), P1s averaged 12.5 item non-responses per 100 surveys. The higher extent of non-response in teacher-reported SDQ may reflect that the majority of teachers

completed the SDQ on paper (for example, 83.8% on paper versus 16.2% online in Wave 10), which meant they were freely able to skip any items.

Across informants, cohorts, and waves, the items with the highest level of non-response were:
1. *Gets on better with adults than with other children* – 3.0 non-responses per 100 surveys
2. *Kind to younger children* – 2.8 non-responses per 100 surveys
3. *Steals from home, school or elsewhere* – 2.5 non-responses per 100 surveys
4. *Thinks things out before acting* – 2.5 non-responses per 100 surveys
5. *Many fears, easily scared* – 1.8 non-responses per 100 surveys

> ***Key points***: The extent of non-response to SDQ items was relatively low among parents or carers in this sample, indicating likely acceptability of the items. The extent of non-response was moderate for teachers and children, indicating that there may be some concerns about interpretability and/or acceptability of some items for these respondents.

Findings from the focus group

In our focus group, RAOs commented that they often had to provide clarifying statements or different wording for the SDQ items. This was particularly common in remote settings and for participants completing the interview in an Aboriginal and/or Torres Strait Islander language, or for participants who completed the interview in English but for whom English was not their first language. Words in the SDQ that the RAOs considered particularly problematic included 'restless', 'depressed', 'tearful', 'solitary', and 'nervous or clingy'. In addition, RAOs stated that they often needed to prompt participants about the response options. This is consistent with previous literature describing issues with the SDQ response options for this population.[5,56] However, the specific words and phrases considered problematic in these three studies (LSIC, WAACHS, and SEARCH) are not the same, demonstrating that there is variation in the cultural appropriateness and interpretability of SDQ items,[5,56] by level of remoteness, language group, or other characteristics.

RAOs indicated that the answers participants gave for SDQ items did not always align with the pre-specified response options of not true, somewhat true, and certainly true. Responses such as 'little bit', were particularly common in areas where English was a second language. RAOs sometimes had to interpret how these answers related to the specified response options.

Adding to concerns about the language used in SDQ, stakeholders identified that some of the behaviours implied to be 'problematic' by the SDQ might be considered appropriate behaviour (not problematic) in different contexts – including with respect to remoteness, gender, and developmental stage. For example, RAOs stated that often being sad or tearful might be expected for children living settings where sorry business is commonly occurring. Other stakeholders identified that behaviours such as being solitary might represent periods of reflection and leadership development. Qualitative work by Williamson et al. identified that child behaviours viewed as 'hyperactivity' in the SDQ might not reflect psychopathology but reflect activeness, a need for more exercise, underlying health conditions, or under-stimulation at school.[5] In many remote settings, Aboriginal and Torres Strait Islander children have high levels of autonomy compared to in more urban settings or to non-Indigenous children. Children behaving autonomously might demonstrate behaviours perceived as 'problematic' by others from outside of this cultural context. A similar concern was identified in a study of Sami (Indigenous) children in Norway; the higher levels of autonomy among Sami compared to non-Indigenous children was hypothesised to contribute to low parent-teacher correlations in perceptions of behavioural problems.[20]

Construct validity

Researchers have previously conducted CFA to test construct validity of the SDQ against the hypothesised five factor structure, including within Aboriginal and/or Torres Strait Islander samples (SEARCH, WAACHS).[50,56] However, given uncertainty about the appropriateness of the structure for this population, we conducted exploratory analysis using PCA to test whether the intended factor structure appeared naturally in the data. To our knowledge, PCA has not previously been conducted for the SDQ in an Aboriginal and/or Torres Strait Islander population. PCA was conducted to explore the factor structure of SDQ data by informant and cohort at Wave 10. Analysis was restricted to the K Cohort to allow for comparison across informants.

For both the P1-reported SDQ and the child-reported SDQ, a low number of items (less than 20% of the 300 possible correlation combinations within each informant) had a strong relationship with each other, that is, an inter-item correlation ≥ 0.3 in absolute value. This suggests low suitability for PCA. Further, two P1 items *often complains of headaches, stomach - aches or sickness* and *gets on better with adults than with other children*; and three child items *I usually share with others*, *I am usually on my own*, and *I take things that are not mine* did not correlate with any other items. Based on the lack of inter-item correlation within the P1-reported SDQ and the child-reported SDQ, these items would ordinarily be excluded from the PCA; however, as the SDQ cannot be modified, these items were retained in the analysis. In the teacher-reported SDQ, there was a higher number of items with strong inter-item correlations (around 45% of correlation combinations had $r \geq 0.3$); most correlations were around $r = 0.5$. This indicated that the teacher-reported SDQ were suitable for PCA.

Results from Bartlett's test of sphericity (degrees of freedom = 300) were significant and the KMO score was above 0.6 for all informants, confirming that the data were adequate to proceed with PCA (Table 4.2).

Table 4.2. Results from the Bartlett test of sphericity and KMO Measure of Sampling Adequacy, by SDQ informant, K Cohort Wave 10

SDQ informant	Bartlett test of sphericity χ² (p-value)	KMO
Parent or carer	3186.91 (<0.001)	0.85
Teacher	2574.86 (<0.001)	0.89
Child	2978.38 (<0.001)	0.85

We anticipated a five-component solution, to align with the five domains of the SDQ. However, analysis did not identify a five-component solution. Across the three methods of determining the number of components for extraction, there were inconsistencies in the number of components to be retained for the informants, but a synthesis of findings indicated use of three to four components for teachers and P1s and three components for children.

An orthogonal and oblique rotated solution were performed. The results were similar so the orthogonal rotation was reported as these results are generally easier to interpret and report.[34] Only items with a component loading of ≥ 0.3 were interpreted. The results from the P1 and teacher three-component and four-component solution were compared. The four-component solution explained a greater amount of variance and yielded more interpretable results in relation to the original SDQ subscales. The current report presents the four-component solution for P1s and teachers and the three-component solution for children.

The P1 four-component solution explained 44.7% of the variance. Table 4.3 presents the component loading after rotation, with loadings < 0.3 omitted from the table. The Prosocial and Hyperactivity subscales loaded onto separate components, although additional items (not part of these subscales) also loaded onto these factors. Three of the five Peer Problems items and Emotional Problems items loaded onto one component, and the final component included three Conduct Problem items. Four items had a loading of < 0.3 for each component; *Often complains of headaches, stomach - aches or sickness*, *Often loses temper*, *Generally well behaved, usually does what adults request*, and *Often volunteers to help others (parents, teachers, other children)*.

Table 4.3. PCA for the parent- or carer-reported SDQ, K Cohort Wave 10 (four-component solution)

	Component			
	1	2	3	4
Emotional Problems subscale				
3. Often complains of headaches, stomach - aches or sickness				
8. Many worries or often seems worried		0.400		
13. Often unhappy, depressed or tearful		0.348		
16. Nervous or clingy in new situations, easily loses confidence			0.367	
24. Many fears, easily scared		0.360		
Conduct Problems subscale				
5. Often loses temper				
7. Generally well behaved, usually does what adults request *(Reverse scored)*				
12. Often fights with other youth or bullies them				0.379
18. Often lies or cheats				0.559
22. Steals from home, school or elsewhere				0.550
Hyperactivity subscale				
2. Restless, overactive, cannot stay still for long			0.338	
10. Constantly fidgeting or squirming			0.396	
15. Easily distracted, concentration wanders			0.373	
21. Thinks things out before acting *(Reverse scored)*			0.356	
25. Good attention span, sees work through to the end *(Reverse scored)*			0.343	
Peer Problems subscale				
6. Would rather be alone than with other youth		0.312		
11. Has at least one good friend *(Reverse scored)*	-0.336			
14. Generally liked by other young people *(Reverse scored)*	-0.392			
19. Picked on or bullied by other young people		0.304		
23. Gets along better with adults than with other young people		0.305		
Prosocial subscale				
1. Considerate of other people's feelings	0.329			
4. Shares readily with other youth, for example books, games, food	0.350			
9. Helpful if someone is hurt, upset or feeling ill	0.350			
17. Kind to younger children	0.333			
20. Often volunteers to help others (parents, teachers, other children)				

The teacher four-component solution explained 59.4% of the variance. Table 4.4 presents the components after rotation, with loadings < 0.3 omitted from the table. All subscales except Peer Problems somewhat loaded onto separate components. Component 2 represented the Hyperactivity subscale and Component 4 represented the Conduct Problems subscale, both having four of the five items loading ≥ 0.3. Component 1 represented the Prosocial subscale and Component 3 represented the Emotional Problems subscale; both included an item from the Peer Problems subscale. Six items had a loading below 0.3 for each component: *Often complains of headaches, stomach - aches or sickness*, *Generally well behaved, usually does what adults request*, *Thinks things out before acting*, *Has at least one good friend*, *Picked on or bullied by other children*, and *Gets along better with adults than with other children*.

Table 4.4. PCA for the teacher-reported SDQ, K Cohort Wave 10 (four-component solution)

	Component			
	1	2	3	4
Emotional Problems subscale				
3. Often complains of headaches, stomach - aches or sickness				
8. Many worries or often seems worried			0.469	
13. Often unhappy, depressed or tearful			0.411	
16. Nervous or clingy in new situations, easily loses confidence			0.366	
24. Many fears, easily scared			0.413	
Conduct Problems subscale				
5. Often loses temper				0.307
7. Generally well behaved, usually does what adults request *(Reverse scored)*				
12. Often fights with other youth or bullies them				0.416
18. Often lies or cheats				0.400
22. Steals from home, school or elsewhere				0.509
Hyperactivity subscale				
2. Restless, overactive, cannot stay still for long		0.378		
10. Constantly fidgeting or squirming		0.401		
15. Easily distracted, concentration wanders		0.399		
21. Thinks things out before acting *(Reverse scored)*				
25. Good attention span, sees work through to the end *(Reverse scored)*		0.349		
Peer Problems subscale				
6. Would rather be alone than with other youth			0.329	
11. Has at least one good friend *(Reverse scored)*				
14. Generally liked by other young people *(Reverse scored)*	-0.358			
19. Picked on or bullied by other young people				
23. Gets along better with adults than with other young people				
Prosocial subscale				
1. Considerate of other people's feelings	0.317			
4. Shares readily with other youth, for example books, games, food	0.394			
9. Helpful if someone is hurt, upset or feeling ill	0.397			
17. Kind to younger children	0.395			
20. Often volunteers to help others (parents, teachers, other children)	0.313			

The child three-component solution explained 40.0% of the variance. Table 4.5 presents the factor loading after rotation, with loadings < 0.3 omitted from the table. The Hyperactivity subscale emerged as a single factor, with four items loading ≥ 0.3. Four Prosocial subscale items loaded onto a component as well as two additional items. The final component included two Emotional Problems items and one Peer Problems item. Twelve items had a loading < 0.3 for each component (Table 4.5).

Overall, analysis did not reveal the intended five-component SDQ structure for any of the informants; it revealed three- and four-component structures. Performance was generally better for the Hyperactivity and Prosocial subscales compared to the other subscales, and for the teacher-reported SDQ compared to child- or P1-reported SDQ, though there was still some deviation from the intended structure. The lack of fit with the hypothesised five-factor structure of the SDQ is not uncommon; within the international literature, there is mixed evidence about the internal factor structure of the SDQ.[49] Goodman has also proposed an alternate structure of a broad Internalising subscale (Emotional Symptoms and Peer Problems combined) and Externalising subscale (Conduct Problems and Hyperactivity combined);[49] however, our findings (the three-factor solution or the four-factor solution) were not consistent with this alternate factor structure.

Across informants, the Peer Problems subscale did not perform as intended, consistent with previous research.[50,56] For all informants, some of the Peer Problems items loaded onto the Prosocial scale. These items were reverse scored, so although the original wording was not in the opposite direction to the prosocial scale, the final responses were opposite. This result was expected as the prosocial items negatively correlated with all items from the remaining four subscales. This fits with previous research identifying a strong correlation between the Peer Problems and Prosocial subscales in an Aboriginal sample. The authors suggest that positive relationships (i.e. the absence of Peer Problems) may be linked to prosocial behaviours and to SEWB; 'It has been noted that in addition to having

strong kinship connections, Aboriginal Australians have strong systems of mutual obligation, which include things like sharing and looking after others. The correlation between peer relationships and prosocial behaviours in the current study appears to reflect this'.[50 p. 168]

Table 4.5. PCA for the child-reported SDQ, K Cohort Wave 10 (three-component solution)

	Component		
	1	**2**	**3**
Emotional Problems subscale			
3. I get a lot of headaches, stomach-aches or sickness			
8. I worry a lot		0.337	
13. I am often unhappy, depressed or tearful		0.361	
16. I am nervous in new situations. I easily lose confidence			
24. I have many fears, I am easily scared			
Conduct Problems subscale			
5. I get very angry and often lose my temper			
7. I usually do as I am told *(Reverse scored)*	-0.304		
12. I fight a lot. I can make other people do what I want			
18. I am often accused of lying or cheating			
22. I take things that are not mine from home, school or elsewhere			
Hyperactivity subscale			
2. I am restless, I cannot stay still for long			0.491
10. I am constantly fidgeting or squirming			0.453
15. I am easily distracted, I find it difficult to concentrate			0.372
21. I think before I do things *(Reverse scored)*			
25. I finish the work I'm doing. My attention is good *(Reverse scored)*			0.311
Peer Problems subscale			
6. I would rather be alone than with people of my age			
11. I have one good friend or more *(Reverse scored)*	-0.324		
14. Other people my age generally like me *(Reverse scored)*			
19. Other children or young people pick on me or bully me			
23. I get on better with adults than with people my own age		0.332	
Prosocial subscale			
1. I try to be nice to other people. I care about their feelings	0.351		
4. I usually share with others, for example CDs, games, food			
9. I am helpful if someone is hurt, upset or feeling ill	0.395		
17. I am kind to younger children	0.348		
20. I often volunteer to help others (parents, teachers, children)	0.321		

Key points: Analysis did not reveal the intended five-component SDQ structure for any of the informants. Analysis was also not consistent with the broader hypothesised Internalising and Externalising subsets of items. This suggests that the SDQ subscales may not tap into the intended constructs in this population; however, this does not necessarily undermine the SDQ's utility in predicting mental health-related conditions in this sample.

Internal consistency

Table 4.6 summarizes the internal consistency findings for the SDQ subscales and Total Difficulties Score; Tables 4.7, 4.8, and 4.9 provide detailed results by wave (age), cohort, informant, level of remoteness, need for translation, and gender. Overall, the internal consistency was acceptable for the SDQ Total Difficulties Score for all three informants and both cohorts at all waves it was administered. In addition, for the Prosocial subscale, internal consistency was acceptable for teacher-reported and child self-reported SDQ at all waves administered, and was acceptable or approaching acceptable for P1-reported SDQ at all but one administration (α = 0.58 for the B Cohort, Wave 6).

Table 4.6. Summary of internal consistency of SDQ subscales and Total Difficulties Score, by SDQ informant and gender, and in very remote areas and among children aged 2-4 years

Scale	Overall: all waves			Females: W10			Males: W10			Very remote: W10			Age 2-4: B Cohort, W3	
	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC
Emotional Problems	~	✓	~	✓	✓	✓	~	✓	~	~	✓	X	X	✓
Conduct Problems	~	✓	~	X	✓	X	~	✓	~	X	✓	X	~	✓
Hyperactivity	~	✓	~	X	✓	~	✓	✓	~	X	✓	X	~	✓
Peer Problems	X	~	X	X	X	X	X	✓	X	X	X	X	X	~
Prosocial	~	✓	✓	✓	✓	✓	✓	✓	~	~	✓	✓	~	✓
Total Difficulties Score	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	~	✓	✓

P1 = parent or carer. TC = teacher. SC = Study Child. X = unacceptable (most alphas <0.6). ~ = approaching acceptable (most alphas 0.6 -< 0.7). ✓ = acceptable (most alphas 0.7 - 0.9).

The Peer Problems subscale was not acceptable for any informant at any wave. The poor performance of this subscale is consistent with previous research in this population.[50,56,63] This is in line with previous research noting that the SDQ's exclusive focus on peers does not capture important relationships, such as with the broader family and community.[47] It may also reflect a lack of conceptual equivalence of the items across cultures; that is, the items, or the behaviours they describe, might not be interpreted or perceived the same way in this population compared to in the non-Indigenous population (or other populations).[64-67]

To test the robustness of findings to different measures of internal consistency, and for consistency with the validation work in WAACHS, analysis was repeated using Raykov's reliability for Wave 10, by informant and cohort. No material differences in the results were observed using Cronbach's alpha compared to Raykov's reliability (data not shown).

Differences by SDQ informant

Patterns of internal consistency were broadly similar for P1-reported and child-reported SDQ for the K Cohort at Wave 10, the first time Study Children self-completed the SDQ. Across waves, cohorts, and subscales, internal consistency was higher for teacher-reported compared to P1-reported or children-reported SDQ.

This could be attributable to the fact that teachers have a larger comparison group than parents or carers (i.e. all of their students in that age group), and/or that they may be more likely to witness peer problems or emotional problems within the school compared to home setting. This hypothesis is supported by the finding that Cronbach's alpha is higher for teacher-reported SDQ compared to P1-reported SDQ in the Longitudinal Study of Australian Children (LSAC) (Table 4.9). However, in most cases, the gap in Cronbach's alpha between the P1-reported SDQ and the teacher-reported SDQ was greater in LSIC compared to in LSAC (Table 4.10). This suggests that other factors (such as Indigenous status, language, and context) might be contributing to the observed differences. For example, in LSIC, all children and the majority of parents or carers identify as Aboriginal and/or Torres Strait Islander, compared to a small minority of teachers. This will be explored in more depth below.

Differences across age and time

Internal consistency was broadly consistent across ages (waves) for each informant, with one exception. Internal consistency for P1-reported SDQ subscales was particularly low in the B Cohort at Wave 3, as low as 0.33 for Peer Problems and 0.41 for emotional problems (Table 4.8). At this wave, 35.8% of children in the B Cohort (n = 295) were aged 2 - < 3 years. The SDQ was originally designed to be administered to children of a minimum age of 3 years. In June 2014, the SDQ early-years version (age 3 - 4 years) was relabelled to include 2-year-olds (2 - 4 years), based on evidence from other populations that the SDQ had good psychometric properties in children aged 2 -< 3 years.[68] Our findings suggest that the SDQ may not be valid for use with Aboriginal and Torres Strait Islander

children at this young age; this could be explored in further research. In general, it is challenging to use any tools to screen for mental health conditions with children this young.[55,69] However, we do note that for this same age group, internal consistency was acceptable for the P1-reported SDQ Total Difficulties Score, and for the teacher-reported Total Difficulties Score and all teacher-reported SDQ subscales except Peer Problems.

Differences by cohort

LSIC employs a cross-sequential cohort design, which enables comparison of two cohorts of children of the same age. For example, we can compare P1-reported and teacher-reported SDQ in two cohorts of children aged 5-7 years: the K Cohort at Wave 3, and the B Cohort at Wave 6. The pattern of results was very similar between the two cohorts. The pattern of results is also very similar for teacher-reported SDQ at age 4-6, the K Cohort at Wave 2 and the B Cohort at Wave 5 (Tables 4.8 and 4.9). This suggests the appropriateness of combining cohorts when the SDQ is administered for children at the same age.

In addition, we can compare the two cohorts of children at the same wave, when they are on average 3 years apart. For example, internal consistency for Wave 10 P1-reported and SDQ-recorded SDQ scales are very similar for the B Cohort (age 9 - 11 years) and the K Cohort (age 12 - 14 years). SDQ scores are intended to be comparable across versions (i.e. ages of children). Accordingly, our findings provide support that cohorts can be combined at the same wave (i.e. when they are three years apart in age). However, analysts may need to consider the difference in cut-off points used to classify SDQ when a different survey age version is used (2-4 years compared to 4-17 years). This occurs in Wave 3 and in Wave 4 (Table 4.1), where the 2-4 year SDQ survey version is used with the B Cohort, and the 4-10 year SDQ survey version is used with the K Cohort. In these instances, the raw SDQ subscale and Total Difficulties Scores may not be comparable between the two cohorts, but the SDQ risk categories are intended to be comparable across versions.

Differences by level of remoteness

We compared the internal consistency of SDQ (all informants) across levels of remoteness at Wave 10, with both cohorts combined (Table 4.11). The internal consistently of P1-reported and child-reported SDQ subscales generally decreased with increasing levels of remoteness, consistent with findings from WAACHS.[56] The only P1- and child-reported SDQ scales with acceptable internal consistency within very remote areas were the Total Difficulties Score (parent or carer) and the Prosocial subscale (child). The observed low alphas may indicate high measurement error, and low inter-relatedness of items for surveys administered in remote and very remote areas. This may be a result of respondents within these settings interpreting the SDQ concepts differently.

However, this same pattern was not observed among teacher-reported SDQ. Internal consistency of teacher-reported SDQ was generally consistent across levels of remoteness. It was 'too high' ($\alpha > 0.90$) for two subscales (Hyperactivity and Prosocial) in the very remote context, and was 'too high' ($\alpha > 0.90$) for the Total Difficulties Score in both the remote and very remote contexts. This might indicate that there is an issue with item redundancy in these contexts.

With increasing level of remoteness, the percentage of parents or carers identifying as Aboriginal and/or Torres Strait Islander increases while the respective percentage for teachers remains relatively consistent and very low (Table 4.7 for data from Wave 10 as an example). These findings seem to indicate that, particularly within remote contexts, the SDQ items are interpreted differently between Indigenous and non-Indigenous respondents.

Table 4.7. Indigenous identification and need for translation by SDQ informant, overall and across levels of remoteness (Wave 10)

	Overall (%)	Major cities (%)	Inner regional (%)	Outer regional (%)	Remote (%)	Very remote (%)
P1 Indigenous	81.4	70.6	75.2	79.4	93.6	100.0
TC Indigenous	4.1	1.7	6.8	4.5	1.8	5.8
Any translation required for P1	6.6	0.0	0.0	10.7	0.0	25.6
Any translation required for SC	4.0	0.0	0.0	3.5	0.7	17.7

P1 = parent or carer. TC = teacher. SC = Study Child.

Differences by respondent need for translation

Zubrick et al. proposed that the reduced internal consistency of SDQ in more remote compared to urban settings in WAACHS might reflect issues in translating the SDQ into local Aboriginal and/or Torres Strait Islander languages.[56] This is consistent with the concerns that the RAOs raised about some of the SDQ items for use in LSIC, as described earlier.

Speaking Aboriginal and/or Torres Strait Islander languages is more common, and accordingly limited English language familiarity is more common, in more remote compared to urban areas. In LSIC, the percentage of P1s and Study Children who required any translation during their interview increased from 0.0% to 25.6% of P1s and from 0.0% to 17.7% of Study Children with increasing level of remoteness (Table 4.7).

We compared the internal consistency of the SDQ subscales and Total Difficulties Score among parents or carers and children who did not need translation compared to those who did need translation (K Cohort, Wave 10). We found that internal consistency for the subscales and the Total Difficulties Score was materially lower among those who required versus those who did not require translation (Table 4.11). We note that numbers are small for the group of parents or carers and children needing translation, and therefore results should be interpreted with caution. However, they provide support for the hypothesis that differences in language use might be contributing to the observed differences in internal consistency by remoteness.

While the number of participants requiring translation was relatively low, this number underestimates the total number of participants for whom English is a second (or later) language, and for whom SDQ language might still be an issue. In any case, it is likely that factors beyond language alone contribute to the observed differences in internal consistency across levels of remoteness. In addition, it is likely that there is variation within level of remoteness (for example, between different mobs or language groups) in the appropriateness of items and internal consistency.

Differences by gender

The internal consistency of SDQ scales was generally similar for male and female study children across informants, when examining the cohorts combined in Wave 10 (Table 4.11). There were two notable differences, where internal consistency was notably lower in female study children compared to males: the P1-reported Hyperactivity subscale and the teacher-reported Peer Problems scale. While internal consistency reached an acceptable level (0.70) for P1-reported hyperactivity in male study children, the alpha was low in females (0.58). Respective figures were 0.71 and 0.49 for the teacher-reported Peer Problems subscale. This was one of few instances in which internal consistency for the Peer Problems subscale was acceptable (for males). In addition, internal consistency for the teacher-reported Total Difficulties Score was high ($\alpha > 0.90$) for male study children, but not for females. This might reflect differences in developmental processes or expectations by gender.

Key points: Internal consistency of the SDQ Total Difficulties Score and the Prosocial subscale was acceptable in nearly all groups and contexts.

Internal consistency was generally greater for teacher-reported SDQ compared to child- or P1-reported SDQ.

Internal consistency was unacceptably low for some subscales for children living in remote and very remote settings. These findings likely reflect differences between SDQ informants (teacher versus P1/child) and differences across levels of remoteness in (A) the understanding of words and concepts in the SDQ, and (B) perspectives of what constitutes 'problematic' behaviour.

Internal consistency was also unacceptably low for children in the B Cohort in Wave 3 of LSIC; the SDQ may not be valid for use at this young age (two to four years) in this population.

When using the SDQ in LSIC, it appears justifiable to combine cohorts, either at the same wave or at the same age. Where different scoring cut-offs apply to the different cohorts based on the SDQ age versions used (2-4 years versus 4-17 years), SDQ categories should be compared rather than raw subscale scores or raw Total Difficulties Score.

Table 4.8. Internal consistency of the SDQ sub-scales and Total Difficulties Score for the B Cohort, by wave and SDQ informant

B COHORT		N*		Emotional Problems		Conduct Problems		Hyperactivity		Peer Problems		Prosocial		Total Difficulties Score	
Wave	Age	P1	TC	P1	TC	P1	TC	P1	TC	P1	TC	P1	TC	P1	TC
1	0-2														
2	1-3														
3	2-4	818	86	0.41	0.85	0.66	0.80	0.63	0.83	0.33	0.62	0.64	0.83	0.73	0.85
4	3-5		206		0.74		0.85		0.89		0.67		0.82		0.87
5	4-6		250		0.75		0.81		0.88		0.66		0.79		0.88
6	5-7	736	300	0.60	0.73	0.62	0.82	0.70	0.90	0.39	0.61	0.58	0.84	0.79	0.88
7	6-8														
8	7-9	752		0.63		0.62		0.69		0.42		0.61		0.80	
9	8-10		355		0.75		0.85		0.88		0.63		0.87		0.89
10	9-11	760	404	0.70	0.76	0.65	0.84	0.73	0.88	0.45	0.69	0.73	0.85	0.83	0.90

*P1 = parent or carer. TC = teacher. Approximate age ranges of Study Children at each Wave are presented in years. *The exact number varies slightly between sub-scales as participants with less than 3 items within a subs-scale were excluded from that analysis. Total Difficulties Score is only calculated for children who have scores calculated for each of the subscales.*

Table 4.9. Internal consistency of the SDQ sub-scales and Total Difficulties Score for the K Cohort, by wave and SDQ informant

K COHORT		N*			Emotional Problems			Conduct Problems			Hyperactivity			Peer Problems			Prosocial			Total Difficulties Score		
Wave	Age	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC
1	3-5																					
2	4-6		114			0.69			0.79			0.90			0.64			0.85			0.88	
3	5-7	578	219		0.57	0.79		0.62	0.80		0.62	0.89		0.40	0.60		0.67	0.82		0.76	0.87	
4	6-8	528	217		0.57	0.76		0.54	0.84		0.68	0.89		0.37	0.63		0.66	0.82		0.75	0.88	
5	7-9		189			0.82			0.87			0.88			0.69			0.86			0.91	
6	8-10	497	205		0.63	0.78		0.68	0.83		0.71	0.89		0.43	0.69		0.67	0.81		0.80	0.88	
7	9-11																					
8	10-12	498			0.61			0.64			0.68			0.43			0.64			0.79		
9	11-13		224			0.76			0.81			0.87			0.68			0.85			0.89	
10	12-14	510	207	484	0.69	0.77	0.69	0.61	0.86	0.60	0.67	0.89	0.66	0.51	0.62	0.49	0.74	0.86	0.71	0.81	0.89	0.81

*P1 = parent or carer. TC = teacher. SC = Study Child. Approximate age ranges of Study Children at each Wave are presented in years. *The exact number varies slightly between sub-scales as participants with less than three items within a sub-scale were excluded from that analysis. Total Difficulties Score is only calculated for children who have scores calculated for each of the subscales.*

Table 4.10. Comparison of internal consistency for the parent- or carer-reported and teacher-reported SDQ in LSAC and LSIC, for children aged 12-14 years

	α			
	LSAC K Cohort, Wave 5		LSIC K Cohort, Wave 10	
Age (years)	12-13		12-14	
SDQ subscale	P1	TC	P1	TC
Emotional problems	0.72	0.74	0.69	0.77
Conduct problems	0.64	0.76	0.61	0.86
Hyperactivity	0.80	0.89	0.67	0.89
Peer problems	0.61	0.68	0.51	0.62
Prosocial	0.71	0.83	0.74	0.86

Analysis of LSAC data conducted by DSS. P1 = parent or carer. TC = teacher.

Table 4.11. Internal consistency of the SDQ sub-scales at Wave 10 for the K Cohort, by level of remoteness and gender

K COHORT	N			Emotional Problems			Conduct Problems			Hyperactivity			Peer Problems			Prosocial			Total Difficulties Score		
Level of remoteness	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC	P1	TC	SC
Major city	140	66	136	0.66	0.74	0.74	0.62	0.83	0.72	0.67	0.83	0.74	0.61	0.52	0.59	0.73	0.83	0.66	0.78	0.85	0.86
Inner Regional	137	58	130	0.73	0.78	0.71	0.69	0.87	0.65	0.77	0.89	0.74	0.64	0.69	0.59	0.79	0.83	0.70	0.86	0.89	0.85
Outer Regional	73	33	69	0.71	0.79	0.75	0.74	0.90	0.48	0.67	0.88	0.68	0.52	0.70	0.44	0.80	0.87	0.69	0.85	0.87	0.80
Remote	54	22	51	0.66	0.81	0.70	0.39	0.86	0.50	0.69	0.86	0.42	0.46	0.73	0.35	0.73	0.82	0.74	0.74	0.92	0.76
Very Remote	106	26	95	0.69	0.74	0.53	0.52	0.84	0.44	0.46	0.92	0.34	0.35	0.33	0.26	0.65	0.94	0.76	0.80	0.92	0.69
SC gender																					
Female	140	66	136	0.71	0.80	0.70	0.59	0.83	0.59	0.58	0.87	0.66	0.49	0.49	0.51	0.75	0.82	0.73	0.80	0.86	0.81
Male	137	58	130	0.68	0.75	0.69	0.63	0.86	0.61	0.70	0.86	0.66	0.52	0.71	0.47	0.72	0.87	0.68	0.82	0.91	0.82
Translation required																					
Any translation required	34		20	0.58		0.54	0.43		0.31	0.53		0.69	0.29		0.40	0.36		0.58	0.67		0.77
No translation required	476		464	0.69		0.70	0.62		0.62	0.69		0.67	0.52		0.50	0.75		0.71	0.82		0.82

P1 = Parent or carer. TC = teacher. SC = Study Child. The exact number varies slightly between sub-scales as participants with less than three items within a sub-scale were excluded from that analysis. Total Difficulties Score is only calculated for children who have scores calculated for each of the subscales.

Convergent validity

The convergent validity of the SDQ Prosocial subscale and Total Difficulties Score was assessed by quantifying the association with outcomes related to mental health. We looked at SDQ scores at Wave 10 for all three informants; the analysis is restricted to the K Cohort to enable comparisons across informants. This analysis is focused on the SDQ Total Difficulties Score and the SDQ Prosocial subscale score; the results of the PCA preclude investigation of the convergent validity of other subscales.

Our primary outcome was P1-report of any child mental health-related conditions in the past 12 months: depression, anxiety, ADD/ADHD, and/or autism. These conditions were measured at Wave 11, one year after SDQ measurement; while data are collected on child autism and ADD/ADHD across waves, Wave 11 is the first time data have been collected on child depression and anxiety. Second, we examined SDQ in relation to child self-reported consideration of suicide, as an indicator of suicide risk and poor SEWB. Third, we examined SDQ in relation to parents or carers concern about their child's development.

Mental health-related conditions

The aim of this analysis was to identify if the scales were performing as intended, i.e. if the SDQ Total Difficulties Score was able to differentiate between children with versus without mental health-related conditions. We compared the prevalence of the mental health outcomes across categories of the SDQ Total Difficulties Score and Prosocial subscale, based on the standard three-band classification system for children aged 4-17 years, and calculated odds ratios (ORs) and 95% Confidence Intervals (CIs) of the child having versus not having the mental health outcome. We note that the prevalence of mental health-related outcomes may be underestimated due to the reliance on P1-report rather than a 'gold standard' measure such as clinical diagnosis.

Higher scores on the Total Difficulties Score are intended to reflect *higher* levels of social and emotional behavioural difficulties. In contrast, higher scores on the Prosocial subscale are intended to reflect *lower* levels of social and emotional behavioural difficulties. Both are categorised as normal, borderline, and abnormal, such that increasing category (Total Difficulties Score and Prosocial subscale) was hypothesised to be associated with increasing prevalence and increasing odds of the mental health outcome. However, we note that the prosocial behaviour does not reflect psychopathology, and is therefore not 'expected' to be a good predictor of mental health diagnoses.[70]

Table 4.12 presents the results according to informant. Overall, the observed pattern of results was consistent with our hypotheses. Despite evidence that the SDQ was not measuring the intended factors in this population, the SDQ Total Difficulties Score was associated with child mental health-related conditions one year later. There was a strong positive association between Total Difficulties Score category and child mental health-related outcomes. While there was some evidence of an association between Prosocial subscale scores and child mental health-related outcomes, the magnitude of the association was weaker compared to the Total Difficulties Score, and most associations were not significant.

The prevalence of any mental health-related condition increased from 3.2% in the 'normal' category to 27.0% to the 'abnormal' category, according to P1-reported SDQ Total Difficulties Score; respective prevalence were 2.2% and 23.3% according to teacher-reported, and 4.8% and 19.6% according to child-reported SDQ. According to P1- and teacher-reported SDQ, odds of a mental health-related condition were more than ten-fold for a child with an 'abnormal' compared to 'normal' Total Difficulties Score, noting wide CIs (P1 OR: 11.23; 95%CI:4.85,26.0; teacher OR: 13.29; 95%CI:3.20,55.16). The effect was weaker for child-reported SDQ (child OR=4.85; 95%CI:1.98,11.86).

For the Prosocial subscale, the prevalence of any mental health-related condition increased from 7.4% in the 'normal' category to 17.4% to the 'abnormal' category, according to P1-reported SDQ; respective prevalence were 2.6% and 15.0% according to teacher-reported SDQ, and 5.7% and 10.6% according to child-reported SDQ. For the P1-reported and child-reported SDQ, an 'abnormal' compared to 'normal' Prosocial subscale score was associated with around a two-fold increased odds of a mental health-related condition, but the association was not significant. For the teacher-reported SDQ, odds of a mental health-related condition were more than six-fold for a child with an 'abnormal' compared to 'normal' Prosocial subscale score, noting wide CIs (teacher OR: 6.53; 95%CI:1.55,27.51).

Our findings provide evidence that higher Prosocial subscale scores (e.g. in the normal compared to borderline or abnormal category) are associated with lower odds of mental health-related conditions. Further, we found that Prosocial subscale scores were positively associated with a measure of resilience (see Chapter 5 for details). This indicates that the Prosocial subscale is capturing children's strengths, and that prosocial behaviours may be an important component of SEWB. This is supported by qualitative research, which found that the behaviours in the Prosocial scale (such as sharing, helping others, and being kind to younger children) 'appear to be of the utmost importance to Aboriginal parents and may be vital in assessing how well Aboriginal children are considered by their families to interact with others and conduct themselves appropriately'.[50 p. 169] However, the Prosocial subscale is only one potential component of SEWB.

Table 4.12. SDQ Total Difficulties Score and Prosocial subscale score at Wave 10 in relation to parent- or carer-reported child mental health-related outcomes at Wave 11, K Cohort, by SDQ informant

	No mental health-related condition	Any mental health-related condition	Odds Ratio of any compared to no mental health-related condition
	% (n)	% (n)	(95%CI)
Total Difficulties Score			
Parent or carer			
Normal (*n* = 314)	96.8 (304)	3.2 (10)	1 (ref)
Borderline (*n* = 47)	85.1 (40)	14.9 (7)	**5.32 (1.92, 14.76)**
Abnormal (*n* = 63)	73.0 (46)	27.0 (17)	**11.23 (4.85, 26.03)**
Teacher			
Normal (*n* = 134)	97.8 (131)	2.2 (3)	1 (ref)
Borderline (*n* = 22)	86.4 (19)	13.6 (3)	**6.89 (1.30, 36.67)**
Abnormal (*n* = 30)	76.7 (23)	23.3 (7)	**13.29 (3.20, 55.16)**
Child			
Normal (*n* = 314)	95.2 (299)	4.8 (15)	1 (ref)
Borderline (*n* = 52)	86.5 (45)	13.5 (7)	**3.10 (1.20, 8.02)**
Abnormal (*n* = 46)	80.4 (37)	19.6 (9)	**4.85 (1.98, 11.86)**
Prosocial subscale			
Parent or carer			
Normal (*n* = 364)	92.6 (337)	7.4 (27)	1 (ref)
Borderline (*n* = 37)	92.0 (34)	8.1 (3)	1.01 (0.32, 3.82)
Abnormal (*n* = 23)	82.6 (19)	17.4 (4)	2.63 (0.83, 8.28)
Teacher			
Normal (*n* = 114)	97.4 (111)	2.6 (3)	1 ref
Borderline (*n* = 33)	87.9 (29)	12.1 (4)	**5.10 (1.08, 24.09)**
Abnormal (*n* = 40)	85.0 (34)	15.0 (6)	**6.53 (1.55, 27.51)**
Self-report			
Normal (*n* = 316)	94.3 (298)	5.7 (18)	1 (ref)
Borderline (*n* = 48)	83.3 (40)	16.7 (8)	**3.31 (1.35, 8.11)**
Abnormal (*n* = 47)	89.36 (42)	10.64 (5)	1.97 (0.70, 5.59)

Any mental health-related condition includes autism, ADD/ADHD, and anxiety and/or depression. A bolded Odds Ratios signifies significant association. Categorised according to three-band classification system for surveys for children aged 4-17 years.

Criterion validity

We then examined the predictive validity of the SDQ in relation to child mental health conditions one year after the SDQ assessment (Table 4.13). As before, SDQ categories were defined using the standard three-band classification system; for this analysis, the 'normal' and 'borderline' categories were combined and compared to the 'abnormal' category, in line with the approach used by Goodman et al.[71]

As shown in Table 4.13, across informants, specificity for the Total Difficulties Score was high, at close to 90%. This means that 90% of all children who did not have a mental health-related condition at Wave 11 were classified as having a 'normal' or 'borderline' Total Difficulties Score at Wave 10. In contrast, sensitivity was low across informants, and far below the desired range of 70-80%; only 30-50% of children with a mental health-related condition were classified as having an 'abnormal' Total Difficulties Score. The PPV of the Total Difficulties Score was 20-30% across informants, meaning that 20-30% of children within the 'abnormal' Total Difficulties Score category at Wave 10 had a mental health-related condition at Wave 11. This means that 70-80% of children within the 'abnormal' Total Difficulties Score category at Wave 10 did not have a mental health-related condition at Wave 11. The low PPV might partly be explained by the low prevalence of the outcome in the sample.

These findings indicate that while there is a strong association between the SDQ Total Difficulties Score and child mental health-related condition, SDQ Total Difficulties Score is a poor predictor of this outcome, according to the cut-offs used. It is likely that the SDQ Total Difficulties Score is capturing experiences of psychological distress, which in turn is associated with mental health-related conditions; these associations are difficult to untangle. Accordingly, Zubrick et al. refer to the SDQ Total Difficulties Score as a measure of 'mental health distress'.[56 p. 22]

For the Prosocial subscale, specificity was high across informants, at 80-90%. This means that 80-90% of all children who did not have a mental health-related condition at Wave 11 were classified as having a 'normal' or 'borderline' Prosocial subscale score at Wave 10. In contrast, sensitivity was very low across informants, and particularly low for the P1-reported (11.8%) and child-reported (16.1%) SDQ Prosocial subscale score; sensitivity was higher for the teacher-reported (46.2%) SDQ Prosocial subscale score, but still far below the desired range of 70-80%. As expected given that prosocial behaviour is not directly tied to psychopathology,[71] the PPV of the SDQ Prosocial subscale score was low, at 10-20% across informants, meaning that 80-90% of children in the 'abnormal' category were misclassified.

Table 4.13. Sensitivity, specificity, and PPV of the SDQ Total Difficulties Score and Prosocial subscale score (dichotomised) in relation to any mental health-related condition, by SDQ informant

SDQ informant	Sensitivity (%)	Specificity (%)	PPV (%)
Total Difficulties Score			
Parent or carer	50.00	88.21	26.98
Teacher	53.85	86.71	23.33
Child	29.03	90.29	19.57
Prosocial subscale			
Parent or carer	11.76	95.13	17.39
Teacher	46.15	80.46	15.00
Child	16.13	88.95	10.64

*According to dichotomised three-band classification system ('abnormal' compared to 'normal/borderline') for surveys for children aged 4-17 years.

Key points: Across SDQ informants, the SDQ Total Difficulties Score had low sensitivity and positive predictive value in relation to child mental health-related conditions, according to current cut-offs. It is likely that the SDQ captures experiences of psychological distress, which in turn is associated with mental health-related conditions. The SDQ Prosocial subscale was not predictive of child mental health-related conditions.

Next we calculated Receiver Operating Characteristics (ROCs) for the Total Difficulties Score. We did not calculate ROC curves for the Prosocial subscale, given that the Prosocial subscale is generally not considered an indicator of psychopathology.[71] We note that the number of children with each condition, according to the parent or carer report, is very low; results should be interpreted with caution.

The ROC curves in Figure 4.1 demonstrate the trade-off between sensitivity and specificity. Across all informants, a lower cut-off point than the boundary for 'abnormal' (in the three-band classification system) would be needed to allow the measure to reach a sensitivity of around 70% (Table 4.14). For the P1-reported Total Difficulties Score, this would require dropping the cut-off point from 17 down to 13; for teacher-reported Total Difficulties Score, this would require dropping the cut-off point from 16 down to 11; for child-reported Total Difficulties Score, this would require dropping the cut-off point from 20 down to 10. These cut-offs correspond to a specificity of 78.0%, 75.7%, and 50.9%, respectively. The proposed cut-offs nearly align with the lower bound of the 'borderline' category for parents or carers (14) and for teachers (12). However, for child-reported SDQ, the proposed cut-off (10) falls within the 'normal' category for children (0-15). At these suggested cut-offs, specificity was substantially lower for child-reported (50.9%) compared to P1-reported (78.0%) and teacher-reported (75.8%) SDQ Total Difficulties Score.

We highlight that the proposed cut-offs are for use in research; these are not applicable for use in clinical settings. Further research would be required to identify clinically-relevant cut-off points.

Table 4.14. Sensitivity and specificity of current and recommended SDQ Total Difficulties Score cut-off point, in relation to parent- or carer-reported child mental health outcomes, by SDQ informant

SDQ informant	Cut-off point for SDQ Total Difficulties Score, for children aged 4-17 years						
	Current*				Recommended		
	Cut-off point	Sensitivity (%)	Specificity (%)		Cut-off point	Sensitivity (%)	Specificity (%)
Parent or carer	≥ 17	50.0	88.2		≥ 13	70.6	78.0
Teacher	≥ 16	53.9	86.7		≥ 11	76.9	75.7
Child	≥ 20	29.0	90.3		≥ 10	71.0	50.9

According to dichotomised three-band classification system ('abnormal' compared to 'normal/borderline') for surveys for children aged 4-17 years. For the three-band classification system, the cut-off point for 'borderline' is 14 for P1-reported SDQ, 12 for teacher-reported SDQ, and 16 for child-reported SDQ.

Key points: Where the SDQ Total Difficulties Score is used in research as an indicator of risk of psychopathology among children in LSIC, lower cut-off points than the 'abnormal' cut-off point are recommended to achieve acceptable sensitivity. Recommendations about cut-offs do not apply to clinical practice. Child-reported compared to P1-reported and teacher-reported SDQ was a poorer predictor of child mental health-related conditions.

Figure 4.1. ROC curve for SDQ Total Difficulties Score at Wave 10 in relation to parent or carer-reported child mental health-related conditions at Wave 11, K Cohort, by SDQ informant
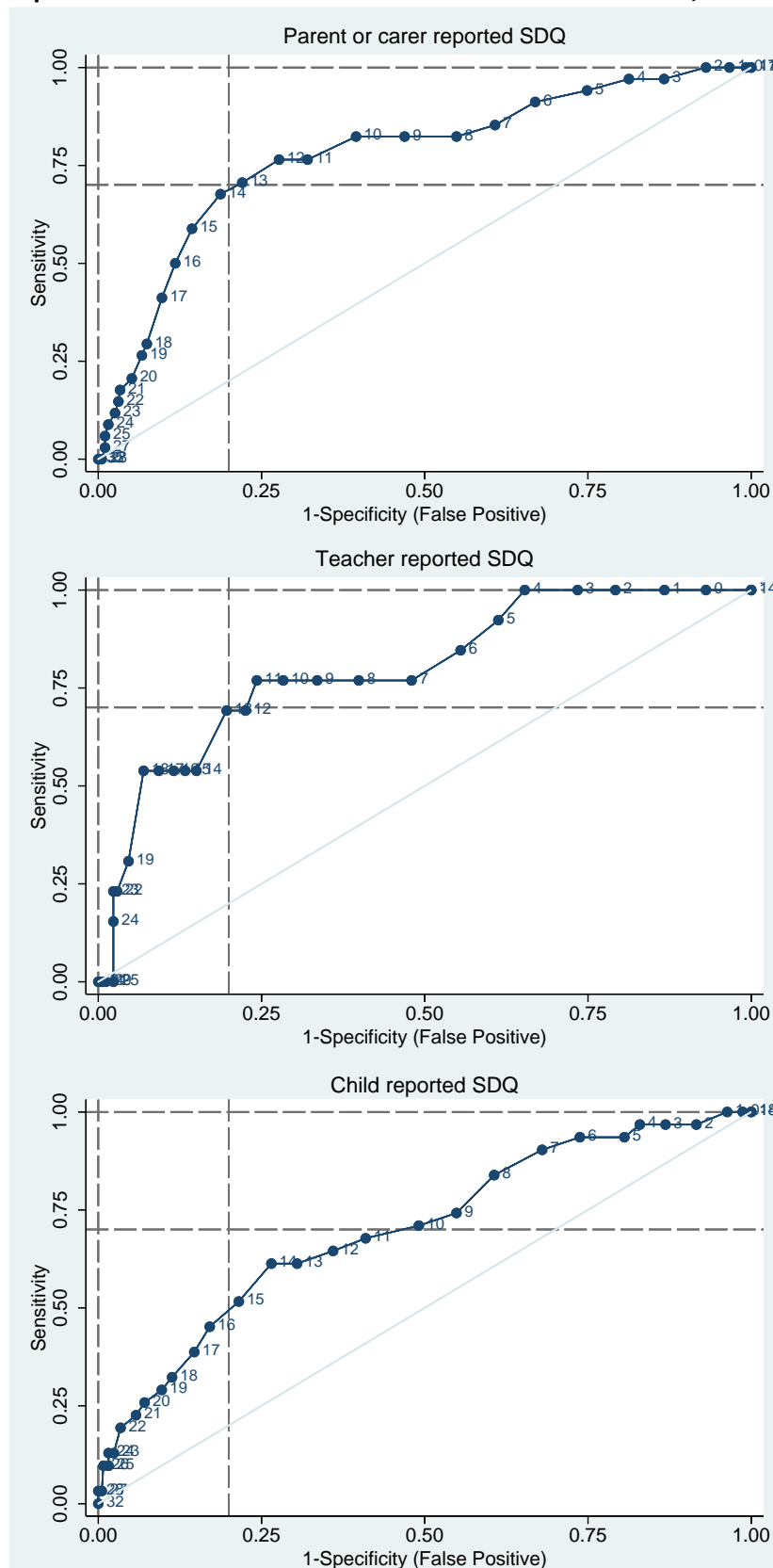
The dotted lines represent the desired minimum sensitivity of 0.70 and specificity of 0.80.

Suicide risk

SDQ is not designed to screen for suicide risk. However, given that this is an outcome of great importance in Aboriginal and Torres Strait Islander communities, and given that it is a marker of poor SEWB, we examined whether the SDQ Total Difficulties Score and the Prosocial subscale score were associated with, and/or predictive of, suicide risk.

In Wave 11, children in the K Cohort were asked to self-complete questions about suicidal thoughts and behaviours, if consent was provided by their parent or carer. Children were asked if they had considered suicide in the last 12 months: *During the past 12 months, did you ever seriously consider attempting suicide?* If they answered yes, children were asked follow-up questions about having planned and attempted suicide, and the number of attempts in the past 12 months. The current analysis is focused on consideration of suicide, as this item was asked of all children whose parent or carer consented.

Table 4.15 presents the association between SDQ scores and child suicide ideation, according to SDQ informant. Overall, the observed pattern of results was consistent with our hypotheses. There was a strong positive association between Total Difficulties Score category and child suicide ideation. There was no evidence of an association between Prosocial subscale scores and child suicide ideation.

The prevalence of child suicide ideation was 3.1% in the 'normal' category, 2.9% in the 'borderline' category, and 11.4% to the 'abnormal' category, according to P1-reported SDQ Total Difficulties Score. Respective prevalence were 3.4%, 5.9%, and 12.5% according to teacher-reported SDQ, and 1.5%, 16.2%, and 13.5% according to child-reported SDQ. The odds of the child considering suicide were around four-fold for children in the 'abnormal' compared to 'normal' category, according to both P1- and teacher-reported SDQ Total Difficulties Score, though the association was not significant for teacher-reported SDQ. The magnitude of association was much stronger for child-reported SDQ, with OR of 12.44 (95%CI:3.33,46.50) for the 'borderline' and 10.04 (95%CI:2.56,39.32) for the 'abnormal', compared to 'normal' category. We suggest caution in interpreting findings due to the large confidence intervals around estimates.

There was not a clear pattern of association between categories of the Prosocial subscale and child suicide risk. For all three SDQ informants, there was not a material difference in the prevalence of child suicide ideation across Prosocial subscale categories, and no associations were statistically significant. This suggests that the SDQ Prosocial subscale score is not associated with child suicide risk in this sample. More research is needed to develop screening tools for suicide risk.

Table 4.15. SDQ Total Difficulties Score and Prosocial subscale score at Wave 10 in relation to child self-reported suicide ideation at Wave 11, K Cohort, by SDQ informant

	Did not consider suicide	Did consider suicide	Odds Ratio of considered compared to did not consider suicide
	% (n)	% (n)	(95%CI)
Total Difficulties Score			
Parent or carer			
Normal (*n* = 314)	96.92 (252)	3.08 (8)	1 (ref)
Borderline (*n* = 47)	97.14 (34)	2.86 (1)	0.93 (0.11, 7.64)
Abnormal (*n* = 63)	88.64 (39)	11.36 (5)	**4.04 (1.26, 12.97)**
Teacher			
Normal (*n* = 134)	96.61 (114)	3.39 (4)	1 (ref)
Borderline (*n* = 22)	94.12 (16)	5.88 (1)	1.78 (0.19, 16.95)
Abnormal (*n* = 30)	87.50 (21)	12.50 (3)	4.07 (0.85, 19.52)
Child			
Normal (*n* = 314)	98.47 (257)	1.53 (4)	1 (ref)
Borderline (*n* = 52)	83.78 (31)	16.22 (6)	**12.44 (3.33, 46.50)**
Abnormal (*n* = 46)	86.49 (32)	13.51 (5)	**10.04 (2.56, 39.32)**
Prosocial subscale			
Parent or carer			
Normal (*n* = 364)	95.97 (286)	4.03 (12)	1 (ref)
Borderline (*n* = 37)	96.00 (24)	4.00 (1)	0.99 (0.12, 7.97)
Abnormal (*n* = 23)	93.75 (15)	6.25 (1)	1.59 (0.19, 13.04)
Teacher			
Normal (*n* = 114)	94.12 (96)	5.88 (6)	1 (ref)
Borderline (*n* = 33)	96.15 (25)	3.85 (1)	0.64 (0.07, 5.56)
Abnormal (*n* = 40)	96.77 (30)	3.23 (1)	0.53 (0.06, 4.60)
Self-report			
Normal (*n* = 316)	95.13 (254)	4.87 (13)	1 (ref)
Borderline (*n* = 48)	97.30 (36)	2.70 (1)	0.54 (0.07, 4.27)
Abnormal (*n* = 47)	51.61 (16)	4.48 (15)	0.65 (0.07, 5.16)

Any mental health-related condition includes autism, ADD/ADHD, and anxiety and/or depression. A bolded Odds Ratios signifies a significant association. Categorised according to three-band classification system for surveys for children aged 4-17 years.

We examined the sensitivity, specificity, and PPV of the SDQ Total Difficulties Score in relation to child suicide ideation (Table 4.16). We calculated these according to standard three-band SDQ categorisation, dichotomised in two different ways: (A) 'abnormal' compared to 'borderline' and 'normal' combined, and (B) 'abnormal' and 'borderline' combined compared to 'normal'. For both classifications, specificity was ≥ 75% for all informants. For P1- and teacher-reported SDQ, sensitivity was low for both classifications (30-50%), meaning that only 30-50% of all children who had considered suicide were categorised as being in the 'high risk' SDQ category. For child-reported SDQ only, sensitivity reached the desired threshold of 70% (73%) when classification (B) was used. PPV was highest in this group (14.9%); this means that 14.9% of children in the 'borderline/abnormal' category had considered suicide 12 months later, but 84.1% had not. While SDQ is not designed to screen for suicide risk, these findings indicate that the child self-reported SDQ may be picking up factors linked to suicide ideation.

We did not calculate sensitivity, specificity, and/or PPV of Prosocial subscale scores in relation to child suicide ideation, given the absence of evidence of association between the Prosocial subscale score and suicide ideation (Table 4.15 above).

Table 4.16. Sensitivity, specificity, and PPV of the SDQ Total Difficulties Score (dichotomised) in relation to child self-reported suicide ideation, by SDQ informant

SDQ informant	Classification (A) 'Abnormal' compared to 'normal/borderline'				Classification (B) 'Abnormal/borderline' combined compared to 'normal'		
	Cut-off	Sensitivity (%)	Specificity (%)	PPV (%)	Cut-off	Sensitivity (%)	Specificity (%)	PPV (%)
Total Difficulties Score								
Parent or carer	≥ 17	35.7	88.0	11.4	≥ 14	42.9	77.5	7.6
Teacher	≥ 16	37.5	86.1	12.5	≥ 12	50.0	75.5	9.8
Child	≥ 20	33.3	90.0	13.5	≥ 16	73.3	88.3	14.9

According to the SDQ three-band classification system, for surveys for children aged 4 - 17 years, dichotomised.

Key points: SDQ is not designed to screen for suicide risk. There is some evidence that SDQ Total Difficulties Score is associated with, but not a strong predictor of, child suicide ideation in LSIC. The association was stronger for child-reported SDQ compared to P1- or teacher-reported SDQ; sensitivity for the child-reported SDQ was acceptable when a lower cut-off point was used, but positive predictive value remained very low. The child self-reported SDQ may be may be picking up factors linked to suicide ideation. We did not find evidence of an association between SDQ Prosocial subscale scores and child suicide ideation.

Parental concern about child behaviour

Researchers have argued that it is flawed to validate measures against other measures that have not been adequately valid and/or may not be relevant to the population of interest, including clinical diagnosis of a mental health condition. For example, Pace et al. write:

> While the DSM-III-R gives reference to making diagnostic adjustments based on cultural relevance, it also structures a framework that presupposes a cultural view of maladjustment that is difficult to apply with diverse peoples and is dependent upon clinical judgment that is in tune with American Indian ways and experiences. With American Indian clients, such understanding is overwhelmingly lacking for most mental health services and professionals … The symptoms and behaviors assessed by the SADS-L and derived from the DSM presuppose a limited view of what is relevant for understanding maladjustment that reflect dominant Euro American cultural values and experiences. Thus, it may not be possible to fully validate the MMPI-2 with American Indians by simply using another majority culture measure.

> … For example, on the one hand, those persons who may appear most bizarre or even schizophrenic to many psychologists may be healers or the leaders who determine when to plant or hunt in tribal cultures and they may not only be functional, but they may live relatively non-distressed lives. That is, one cannot assume that the "unusual" thinking captured by the MMPI-2 or the SADS-L will necessarily be linked to distress or to a specific type of meaningful diagnosis in a tribal context. On the other hand, contemporary bio-psycho-social-cultural-historical-political-linguistic contexts and tribal peoples interpretations may have impacted the participants' protocols. For example, a person who is experiencing job instability or rebellion against authority figures may not be dysfunctional, maladjusted, or antisocial, but rather be experiencing real consequences of historical oppression, current racism, and poverty. Without considering these possible influences, personality test results may be misinterpreted.[21 p. 328]

These views are supported by stakeholder input in the current project. Stakeholders commented that the experience/expression of mental health conditions, and a family and community's response to persons with mental health conditions, can vary, including between more remote and more urban settings. In remote contexts, people with mental health conditions are allowed to persist with that

condition, and the family and community will work around the condition rather than trying to 'fix' it. In these contexts, mental health conditions themselves are not necessarily considered problematic.

Accordingly, we aimed to examine the SDQ in relation to an outcome defined by the participants themselves, rather than relying on measures defined according to another population's viewpoint.

The SDQ impact supplement, or the items from PEDS, present an opportunity to focus on outcomes of concern to Aboriginal and Torres Strait Islander peoples. These questions directly ask if the parent or carer has any concerns about their child's development and interactions. These items likely capture distress/behaviours/conditions that parents or carers perceive are impacting on the child's functioning. This measure may miss instances of poor child wellbeing, in particular in relation to internalising disorders which may be harder to recognise from the outside. Figure 4.2 presents P1-reported concern about their child's behaviour and social interactions by level of remoteness, based on responses to two items from PEDS. Consistent with stakeholder input, we observe that parents or carers living in more remote compared to urban areas tend to be less likely to report concerns about their child's behaviour or social interactions.

Figure 4.2. Parent- or carer-reported concern about their child's behaviour and social interactions by level of remoteness (B Cohort, Wave 6)
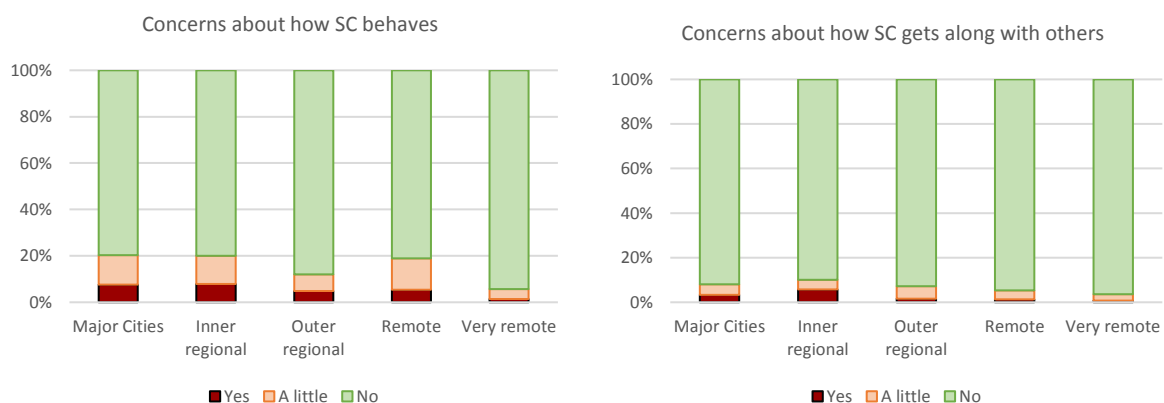
To conduct a preliminary assessment of whether P1-reported concerns were linked to SDQ, we compared the mean P1-reported SDQ Prosocial subscale scores and Total Difficulties Score across levels of parent or carer concern. Overall, results were as expected, with decreasing mean Prosocial subscale score and increasing mean Total Difficulties Score with increasing level of parent or carer concern (Figure 4.3). When we examined this by level of remoteness, we found that the link between P1 concern and P1-reported SDQ scores was weaker in more remote compared to more urban areas (noting small numbers; Figure 4.4). This may suggest that in more remote compared to more urban areas, SDQ scores are less likely to be reflective of behaviours that are of concern to parents or carers. These preliminary analyses support stakeholder views that behaviours and/or mental health-related conditions are viewed differently in remote compared to other contexts.

Key points: These findings support stakeholder views that the experience and expression of mental health conditions varies between more remote and more urban settings.

35

Figure 4.3. Mean SDQ Prosocial subscale score and Total Difficulties Score by level of parent or carer concern about how the SC behaves and how the SC gets along with others ('gets along')
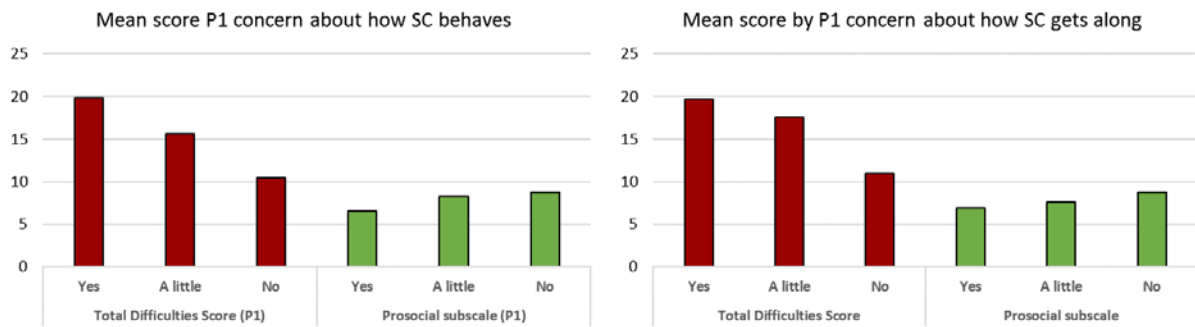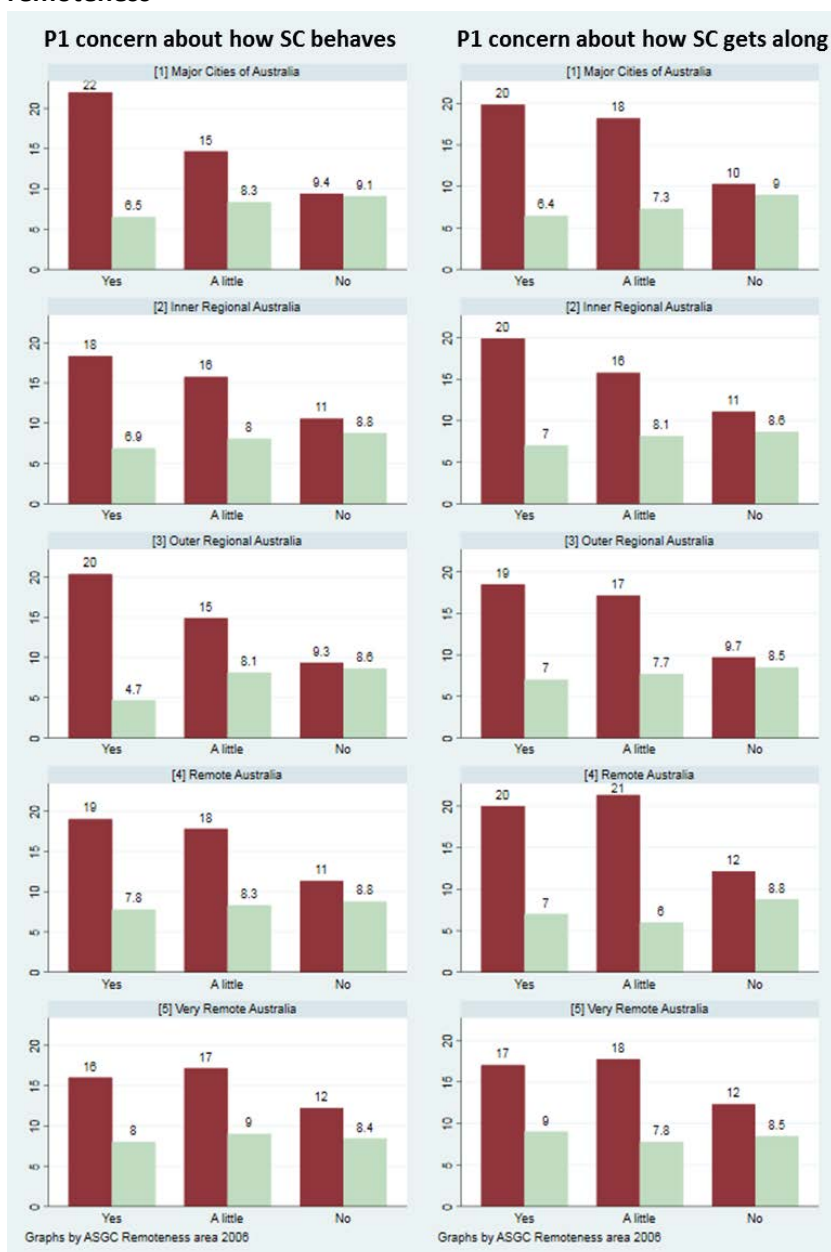

Figure 4.4. Mean SDQ Prosocial subscale score and Total Difficulties Score by level of parent or carer concern about how the SC behaves and how the SC gets along with others ('gets along'), by level of remoteness

Between-informant correlation for the Total Difficulties Score and the Prosocial subscale

We examined correlation in the Total Difficulties Score and the Prosocial subscale between the three SDQ informant groups (Table 4.17): parents or carers, teachers, and children. Overall, correlations between SDQ informants were higher for the Total Difficulties Score (r = 0.42 - 0.45) than the Prosocial subscale score (r = 0.27 - 0.34). Correlations for each outcome were relatively similar across informant pairs; for example the correlation between parent or carer and teacher Total Difficulties Score was 0.42, and the correlation was 0.45 for both parent or carer-children and teachers-children. The observed correlations are of a similar magnitude to the mean cross-informant SDQ correlation in a study of rural Australian children aged 5-16 years (r = .36).[72] This fits with international evidence on the relatively low correlation between parent, teacher, and self-reported child psychopathology, and differences in concordance by children's age, setting, and type of psychopathology (e.g. internalising versus externalising behaviours).

Table 4.17. Pairwise correlation between parent or carer-, teacher-, and child self-reported SDQ scores: Total Difficulties Score and Prosocial subscale

	Total Difficulties Score r	Prosocial subscale r
Parent or carer- teacher	0.42	0.27
Parent or carer - child	0.45	0.33
Teacher - child	0.45	0.34

Key points: There was a weak to moderate level of correlation between SDQ informants in the Total Difficulties Score. The observed level of correlation is consistent with what has been observed in other populations.

Recommendations for data users

1. **Validity for use in the Aboriginal and/or Torres Strait Islander context:** The weight of the evidence to date supports use of the SDQ Total Difficulties Score as a measure of psychological distress – not as a measure of mental disorder or SEWB – and our analyses support using this broad interpretation of the measure among Aboriginal and Torres Strait Islander children in LSIC.
   1.1. However, we have identified substantial limitations to use of the SDQ. The language used and behaviours considered 'problematic' in the SDQ Total Difficulties Score do not necessarily align with perspectives among Aboriginal and Torres Strait Islander peoples, particularly in remote and very remote settings. Further qualitative item validation research across contexts and language groups is required to accept the validity of the SDQ for use with Aboriginal and Torres Strait Islander children.[19,21]
   1.2. Accordingly, SDQ data with this population must be interpreted with caution, and should be interpreted within context.
   1.3. Researchers using the SDQ need to be explicit about the limitations of using this measure with Aboriginal and Torres Strait Islander peoples.
   1.4. It is not appropriate to compare the SDQ between Aboriginal and Torres Strait Islander and non-Indigenous children; we cannot assume that the SDQ is conceptually equivalent (measuring the same thing) in these populations.

2. **Total Difficulties Score as a measure of psychological distress:**
   1.1. SDQ is not a diagnostic instrument, and should not be used to diagnose children with mental health-related conditions.[70] We found that SDQ Total Difficulties Score was associated with, but not a strong predictor of, child mental health-related conditions one year later. It is likely that the Total Difficulties Score is measuring psychological distress, which in turn contributes to mental health-related conditions.
   1.2. If the SDQ is examined in relation to child psychopathology in the research (not clinical) context, we recommend use of P1-reported and teacher-reported SDQ over child-reported SDQ. Where these are examined as binary variables, we recommend a cut-off of ≥ 14 for P1-reported SDQ and ≥ 12 for teacher-reported SDQ (for surveys for children aged 4-17 years).
   1.3. The SDQ is not designed to screen for suicide risk. The SDQ Total Difficulties Score is associated with, but not a strong predictor of, child suicide ideation in LSIC. It is likely that the SDQ is picking up factors linked to suicide ideation.
   1.4. If the SDQ is examined in relation to child suicide risk in the research (not clinical) context, we recommend use of child-reported SDQ. Where this is examined as a binary variable, we recommend a cut-off of ≥ 16 for child-reported SDQ (for surveys for children aged 4-17 years).

3. **Strengths-based use is warranted:** The SDQ Prosocial subscale demonstrated good psychometric properties in this sample, overall and across groups. We recommend use of the Prosocial subscale as a strengths-based measure and component of SEWB. The evidence does not support use of the other subscales on their own.

4. **Validity for younger children is weak:** We recommend not using the SDQ for children in the B Cohort in Wave 3 of LSIC, as the SDQ may not be valid for use at this young age (2-4 years) in this population.

5. **Methods for combining age groups:** When using the SDQ in LSIC, it appears justifiable to combine cohorts, either at the same wave or at the same age. Where different scoring cut-offs apply to the different cohorts based on the SDQ age versions used (2-4 years versus 4-17 years), SDQ categories should be compared rather than raw subscale scores or raw Total Difficulties Score.

5. Strong Souls: Strengths (Resilience)

Background

Purpose of the Strong Souls Index

The Strong Souls Index was developed as a tool to assess the social and emotional well-being (SEWB) of Aboriginal and Torres Strait Islander adolescents aged 16-21 years participating in the longitudinal Aboriginal Birth Cohort (ABC) Study in the Northern Territory.[73] It was designed to include concepts relating to Depression, Anxiety, Suicide Risk, and Resilience. A pool of potential items was created and then refined through a review of the literature and ongoing consultative processes with Aboriginal community members and mental health experts.[73,74] The consultation process also resulted in the tool being designed as a self-complete questionnaire and including four response categories for all items. Through this process the tool was named Strong Souls, 'in recognition that the concept of "soul" encompasses a person's physical, emotional, social and spiritual being and was therefore synonymous with SEWB'.[73 p. 41-2]

A pool of 34 Strong Souls items were tested in a pilot study with 361 Aboriginal participants in the Northern Territory, alongside two existing measures, the K6 and the WASC-Y.[73] The sample included cultural and linguistic diversity, comprising participants from a range of communities, the majority of whom had English as a non-primary language. Analysis was undertaken using data from this pilot sample to refine the Strong Souls Index.

The final Strong Souls Index includes 25 items from the original pool of 34 items, across four factors: Anxiety (6 items), Resilience (9 items), Depression (7 items), and Suicide Risk (3 items).[73] It encompasses two subscales: SEWB (distress) and Resilience.[73] Resilience has been defined as:

> In the context of exposure to significant adversity…. both the capacity of individuals to navigate their way to the psychological, social, cultural, and physical resources that sustain their well-being, and their capacity individually and collectively to negotiate for these resources to be provided in culturally meaningful ways.[75 p. 255]

In the pilot study, the Strong Souls Index demonstrated construct validity, reliability, acceptability, and cultural appropriateness for use as a measure of SEWB with young Aboriginal people in the Northern Territory.[73] To our knowledge, no other studies have evaluated the validity of the Strong Souls Index in another sample. Therefore, it is unknown if these findings of validity are generalisable beyond this age group or geographic area.

The Strong Souls Index has been used as a measure of SEWB in other research.[8,76-79] There is no established scoring guide or cut-offs for the Index; however, in previous research, scale scores have been created by summing items for each factor.[73] An overall SEWB (Distress) score can also be calculated by summing scores across the Anxiety, Depression, and Suicide Risk factors.[8]

Modifications between original scale and scale used in LSIC

Items from the Strong Souls Index have been used with parents or carers in LSIC across waves (for details, see *Technical report: Measuring adult mental health, psychological distress, and social and emotional wellbeing in the Longitudinal Study of Indigenous Children*). Items from the Strong Souls Index are used with LSIC Study Children for the first time in Wave 9. Questions are asked of the K Cohort only, who are around age 11-13 years. The Strong Souls Index was designed for use with adolescents (aged 16-21 years); to our knowledge, an age range for use is not specified.

Full details on the selection of Strong Souls items for use with LSIC parents or carers are provided in the *Technical report: Measuring adult mental health, psychological distress, and social and emotional*

wellbeing in the Longitudinal Study of Indigenous Children. Briefly, some of the Strong Souls items from the initial pool of 34 were selected for inclusion in LSIC, with minor modifications and additional clarifying text in some instances. These items were selected before the Strong Souls validation work was completed, and therefore the items used in LSIC do not reflect the final Strong Souls Index. For parents or carers, LSIC includes a set of items about distress (called Strong Souls: Distress here), and a set of items about Resilience (called Strong Souls: Strengths here). The Strong Souls: Distress items are not asked of the Study Children. The Strong Souls: Strengths items used with Study Children are based on those used with parents or carers in previous waves. There are some changes to the wording and clarifying text used with children compared to with parents or carers in LSIC.

Of the 12 total Strong Souls: Strengths items used in LSIC, 8 are from the final Strong Souls Index, all of which map on to the Resilience factor. This covers 8 of the 9 total items from the Resilience subscale of the Strong Souls Index. The only item from the Resilience subscale that is not included in LSIC (for Study Children at Wave 9) is the item, *You know lots about white fella way*s. This item was asked of Aboriginal and/or Torres Strait Islander parents or carers in earlier waves of LSIC. The item was removed from the survey in response to RAO feedback: '… the item about knowledge of "white fella ways" was dropped partly due to its low loading on the resilience factor in the original study but also due to feedback from the Indigenous researchers that the item was being interpreted differently in different locations and that it was not appropriate to ask this item of non-Indigenous parents of Indigenous children.'[74 p. 7] Because this item was dropped from the parent or carer survey at an earlier Wave, it was not used with Study Children in Wave 9. Two of the remaining four items in the LSIC Strong Souls: Strengths measure are from the initial pool of Strong Souls items, but these two items were not included in the final Strong Souls Index (items 26 and 27 in Table 5.1). The other two items from the LSIC Strong Souls: Strengths measure came from an early stage of Strong Souls development (preliminary items 1 and 2 in Table 5.1). These two items were part of the Strong Souls pool of potential items at the time the measure was provided to LSIC for incorporation into the questionnaire; however, these preliminary items were not in the pool of 34 items assessed in the Strong Souls validation study. The Resilience items from the Strong Souls Index, the Strong Souls items used in LSIC, and their respective response options, are displayed in Table 5.1.

There is an important difference in the way LSIC has coded the response options for the Strong Souls: Strengths items, compared to how they are coded in the Strong Souls Index. As shown in Table 5.1, in LSIC, the response options are coded such that a lower score corresponds to a higher level of agreement with each item. This means that in LSIC, under the default coding system, a lower total score represents a higher level of Resilience. In contrast, in the Strong Souls Index, the response options are coded such that a lower score corresponds to a lower level of agreement (or higher level of disagreement) with each item. This means that in the Strong Souls Index, a lower total score represents a lower level of Resilience.[73] In this Report, to align with the Strong Souls Index, the Strong Souls: Strengths (Resilience) items in LSIC were reverse coded such that a lower score indicates a lower level of Resilience.

Table 5.1. Comparison of the Strong Souls Index final Resilience items against the Strong Souls: Strengths items used in LSIC

Strong Souls Index: Resilience items			LSIC Strong Souls: Strengths items	
Prompt: How much is this like you?			*Prompt: The next questions are about what helps you get through hard times. How much is this like you?*	**Response options**
Item	**Response options**	**Factor**		
From final Strong Souls Index				
17. You have a strong family who help each other.	Always, most times, sometimes, not really	Resilience	B. You have a strong family who help each other.	Always (1), most times (2), sometimes (3), not really (4)
18. You know lots about white fella ways.	Lots, fair bit, little bit, not much	Resilience	--	--
19. You know someone who is a really good person.	Lots of people, fair few, not many, no one	Resilience	J. You know someone who is a really good person.	Lots of people (1), fair few (2), not many (3), no one (4)
20. You laugh and make jokes a lot.	Lots, fair bit, little bit, not much	Resilience	D. You laugh and make jokes a lot.	Lots (1), fair bit (2), little bit (3), not much (4)
21. You are really into something (like music, cars, clothes, football, fishing, computers, etc.).	Lots, fair bit, little bit, not much	Resilience	E. You are really into something (like music, football, clothes)? **Is there something you really like?**	Lots (1), fair bit (2), little bit (3), not much (4)
22. You are a good son or daughter to your family.	Always, most times, sometimes, not really	Resilience	G. You are a good (son or daughter) to your family. **You look after your family.**	Always (1), most times (2), sometimes (3), not really (4)
23. You got an older person looking out for you.	Always, most times, sometimes, not really	Resilience	I. You got an older person looking out for you.	Always (1), most times (2), sometimes (3), not really (4)
24. You got lots of friends.	Lots, fair few, not many, none	Resilience	K. You **have** lots of friends.	Lots (1), fair few (2), not many (3), none (4)
25. When you are upset, you can usually talk to someone about it (parents or friends).	Always, most times, sometimes, not really	Resilience	L. When you're sad **you have a person you can talk to.**	**Always (1), fair bit (2), little bit (3), never (4)**
Not in final Strong Souls Index				
[26.] You get used to big changes in your life quickly		Resilience	C. You get used to big changes **fast. Like changing schools or moving house.**	Always (1), most times (2), sometimes (3), not really (4)
[27.] People say you are really good at something		Resilience	H. People say that you are really good at something. **Sports or fishing or looking after kids.**	Always (1), most times (2), sometimes (3), not really (4)
[34.] You wish you were a different person		Resilience		
[Preliminary 1.*] When you get upset, you can find something to cheer you up. **You got something you can do to make you feel better (ask what works best).**		Resilience*	A. When you get sad **you can** find something that makes you **happy.**	Always (1), most times (2), sometimes (3), not really (4)

[Preliminary 2.*] You know a lot about your Aboriginal culture.		Resilience*	F. You know a lot about your (Aboriginal/ Torres Strait Islander) family history and culture **(e.g. family stories and relationships)**.	Lots (1), fair bit (2), little bit (3), not much (4)

*Bolded text indicates modified text or clarifying statements included in LSIC. * Preliminary Strong Souls items that were provided to LSIC for incorporation into the questionnaire; these preliminary items were not in the pool of 34 items assessed in the Strong Souls validation study.*

Results: assessment of Strong Souls: Strengths in LSIC

Acceptability

Four children were not asked any Strong Souls Resilience item by the RAO's. The extent of non-response was relatively low, averaging 13.8 item non-responses per 100 surveys. The item *You get used to big changes fast* had the greatest rate of non-response, averaging 3.9 item non-responses per 100 surveys.

> **Key points**: Overall, the extent of non-response to Strong Souls items was relatively low in this sample, indicating likely acceptability of the items.

Construct validity

Overall, inter-item correlations were low. Of the 66 possible inter-item correlations, only 14 were above the 0.3 threshold. No inter-item correlation was ≥ 0.4. However, results from the Bartlett's test of sphericity were significant (degrees of freedom = 66) and the KMO score of 0.83 indicated that the data were suitable for PCA.

The three extraction techniques used suggested a two-component solution (parallel analysis and scree plot) and a three-component solution (Kaiser's criterion). Both orthogonal and oblique rotated solutions were assessed. The results were similar so the orthogonal rotation was reported. Only items with a component loading ≥ 0.3 were interpreted.

The two-component solution explained 40.6% of the variance (Table 5.2). Four items had a loading less than the 0.3 threshold *You get used to big changes fast*, *You know someone who is a really good person*, *When you get sad you can find something that makes you happy*, and *People say that you are really good at something*.

Similar to findings from the Strong Souls: Strengths (Resilience) items with parents or carers in LSIC, for the two-component solution, (1) the items that loaded above 0.3 on Component 1 appeared to relate to social support, and (2) the authors did not identify a clear construct underlying the items that loaded above 0.3 on Component 2. Four items did not load above the 0.3 threshold in the two-component solution: *You get used to big changes fast*, *When you get sad you can find something that makes you happy*, and *You know someone who is a really good person*.

Table 5.2. PCA for the Strong Souls: Strengths (Resilience) measure, K Cohort Wave 9 (two-component solution)

	Component	
	1	2
2. You have a strong family who help each other	0.485	
7. You are a good (son or daughter) to your family	0.425	
10. You got an older person looking out for you	0.466	
12. When you're sad you have a person you can talk to	0.401	
5. You laugh and make jokes a lot		0.534
6. You are really into something (like music, football, clothes)?		0.500
8. You know a lot about your (Aboriginal/Torres Strait Islander) family history and culture (e.g. family stories and relationships)		0.317
11. You got lots of friends		0.397
1. When you get sad you can find something that makes you happy		
3. You get used to big changes fast		
4. You know someone who is a really good person		
9. People say that you are really good at something		

There was some consistency between the two- and three-component solutions. The three-component solution explained 49.5% of the variance (Table 5.3). Items that loaded above 0.3 on Component 1 appeared to broadly relate to social support, although there were some differences between items that loaded onto this factor in the two- versus three-component solution. As occurred in the two-component solution, the authors did not identify a clear construct underlying the items that loaded onto Components 2 and 3. As per the two-component solution, the item *You get used to big changes fast* did not load above the 0.3 threshold in the three-component solution. This item had the highest extent of non-response in this cohort, and was not included in the final Strong Souls Index. Together, this may suggest that this item does not function as originally intended. In addition, the item about family history and culture cross-loaded onto Components 1 and 2.

Table 5.3. PCA for the Strong Souls: Strengths (Resilience) measure, K Cohort Wave 9 (three-component solution)

	Component		
	1	**2**	**3**
2. You have a strong family who help each other	0.416		
7. You are a good (son or daughter) to your family	0.565		
9. People say that you are really good at something	0.385		
10. You got an older person looking out for you	0.466		
8. You know a lot about your (Aboriginal/Torres Strait Islander) family history and culture (e.g. family stories and relationships)	0.341	0.330	
5. You laugh and make jokes a lot		0.546	
6. You are really into something (like music, football, clothes)?		0.500	
11. You got lots of friends		0.388	
1. When you get sad you can find something that makes you happy			0.409
4. You know someone who is a really good person			0.439
12. When you're sad you have a person you can talk to			0.632
3. You get used to big changes fast			

Next, the eight items from the final Strong Souls Resilience subscale were tested in a one-component PCA. The solution explained 32.4% of the variance (Table 5.4). All items had a loading ≥ 0.3 except the item *You laugh and make jokes a lot*, which had a loading of 0.298. All eight items were retained and used in the following analyses, given that these eight items were all included in the final Strong Souls Index.

Table 5.4. PCA for the eight-item Strong Souls: Strengths (Resilience) measure, K Cohort, Wave 9 (one-component solution)

	Component
	1
2. You have a strong family who help each other	0.374
4. You know someone who is a really good person	0.386
5. You laugh and make jokes a lot	
6. You are really into something (like music, football, clothes)?	0.317
7. You are a good (son or daughter) to your family	0.395
10. You got an older person looking out for you	0.356
11. You got lots of friends	0.319
12. When you're sad you have a person you can talk to	0.371

> ***Key points***: The Strong Souls: Strengths items used in LSIC are intended to measure a single construct, Resilience. However, our analysis indicated a two- or three-component solution; this suggests that the 12 items used in LSIC had poor construct validity. Our findings support the use of an eight-item Resilience measure, which is comprised of only items from the final Strong Souls Resilience subscale.

Internal consistency

The eight-item Strong Souls Resilience measure had an internal consistency score of $\alpha = 0.69$, which was approaching acceptable ($\alpha \geq 0.70$).

> ***Key points***: Internal consistency approached acceptable for the reduced eight-item Strong Souls Resilience measure.

Convergent and divergent validity

The eight-item Strong Souls Resilience measure was computed by summing scores across all eight items. Children missing responses to one or more items had a missing total score. Of 516 children who completed any Strong Souls items at Wave 9, 414 (80.2%) had an eight-item Strong Souls Resilience score.

To test convergent validity, we quantified the correlation between the eight-item Strong Souls Resilience score at Wave 9, and the child-reported SDQ Prosocial subscale score at Wave 10. We expected a weak or moderate positive correlation, as both are strengths-based measures, but are distinct constructs (resilience versus prosocial behaviour). To test divergent validity, we quantified the correlation between the eight-item Strong Souls Resilience score at Wave 9 and the child-reported SDQ Total Difficulties Score at Wave 10. We expected a weak or moderate negative correlation, as one is a measure of strengths (resilience) and the other is a measure of distress (Total Difficulties Score), and they are measuring distinct constructs. Pearson's r correlation was used.

As hypothesised, there was a weak positive correlation between the eight-item Strong Souls Resilience and the SDQ Prosocial subscale, supporting convergent validity with another component of SEWB measured in LSIC (Table 5.5). Supporting divergent validity of the measure, we observed a weak negative correlation between the eight-item Strong Souls Resilience measure and the SDQ Total Difficulties Score.

Table 5.5. Correlation of the eight-item Strong Souls Resilience measure at Wave 9 (K Cohort) and child-reported SDQ Prosocial subscale score and SDQ Total Difficulties Score at Wave 10

	Pearson's r
SDQ Prosocial subscale score	0.28
SDQ Total Difficulties Score	-0.20

> ***Key points***: The eight-item Strong Souls Resilience measure demonstrated convergent and divergent validity.

Recommendations for data users

1. **Evidence of validity:** The eight-item Strong Souls Resilience measure appeared to be valid, reliable measure of resilience, a component of SEWB. The Strong Souls Index is one of few measures of SEWB developed with and specifically for Aboriginal and Torres Strait Islander peoples.
    1.1. The Strong Souls Index was developed with a diversity of communities, and it performed well in the diverse LSIC sample; further research across contexts and language groups would reinforce the validity of the eight-item Strong Souls Resilience measure.
    1.2. The remaining four items from the Strong Souls: Strengths scale could be used individually or as a group.

2. **Scoring the measure:** We recommend that data users reverse score the Strong Souls: Strengths (Resilience) items in LSIC, for consistency with scoring of the Strong Souls Index.

6. Social Skills Improvement System (SSIS)

Background

Purpose of SSIS

There is international evidence that children with higher compared to lower levels of social skills have more positive academic outcomes, social outcomes, and mental health outcomes (including a lower risk of conduct problems, anxiety, and depression).[42,80-82] The Social Skills Rating System (SSRS), later adapted into the Social Skills Improvement System (SSIS), is a multi-informant screening instrument to assess children's Social Behaviours, Social Skills and Academic Competence.[83-85] The aim of this measure is to generate a holistic assessment of the student's social behaviour across five social skills domains: Cooperation, Assertion, Responsibility, Empathy, and Self-Control; and three problem behaviour domains: Internalizing, Externalizing, and Hyperactivity.[83-85] SSRS includes three forms, to be completed by the student, their parent, and their teacher, to enable collection of complimentary information about the student within the home, school, and community environments.[80] The use of multiple informants is intended to provide different perspectives on the child's behaviour. For example, the Student Form provides insight into the perceptions and beliefs of the child that might not be known by other informants. The SSRS and SSIS can be used with children aged 3-18 years.[84] The SSRS is designed to be administered by psychologists, teachers, learning specialists, and school counsellors who have a basic understanding of the behaviours and cognitive treatments.[85]

The SSRS was replaced by SSIS – Rating Scale (SSIS-RS) in 2008.[84,86] The SSIS-RS was developed to incorporate: 'updated national norms, better measures for children aged 3-5, four additional subscales, generate overlap of content across forms, improved psychometric properties, validity scales, Spanish versions of the Parent and Student Forms, newly developed scoring and reporting software and direct link to interventions'.[87 p. 1] The four new domains include two social skills domains (Communication and Engagement), and two problem domains (Bullying, Autism Spectrum).[84,87] The SSIS-RS has been developed to be administered in a short time period and to be easily readable.[86,87] The SSIS-RS can be completed by hand or on the computer.

The social skills domains assessed within the refined SSIS-RS are defined as:
> **Communication**: taking turns and making eye contact during a conversation, using appropriate tone of voice and gestures, and being polite by saying "thank you" and "Please";
> **Cooperation:** helping others, sharing materials, and complying with rules and directions;
> **Assertion:** initiation behaviours, such as asking others for information, introducing oneself, and responding to the actions of others
> **Responsibility:** showing regard for property or work and demonstrating the ability to communicate with adults
> **Empathy:** showing concern and respect for others' feelings and viewpoints
> **Engagement:** joining activities in progress and inviting others to join in, initiating conversations, making friends, and interacting well with others
> **Self-control:** responding appropriately in conflict (e.g., disagreeing, teasing) and non-conflict situations (taking turns and compromising).[87 p. 1-2]

The problem domains assessed in the SSIS-RS are: Externalizing, Bullying, Hyperactivity/Inattention, Internalizing, and Autism Spectrum.

The SSIS-RS includes different forms for completion by students (different versions for children 8-12 years and 13-18 years), parents, and teachers. All forms assess students' Social Skills and Problem Behaviours; in addition, the Teacher Form assesses Academic Competence.[84,87] The Student Form includes 46 items, the Parent Form 46, and the Teacher Form 30. The SSIS-RS asks Parents and

Teachers to think about the child's behaviour in the past two months and indicate how often they have displayed each behaviour. The student Forms asks students to *decide how true each sentence is for you*, with no specific timeframe attached. The response options used in SSIS-RS are different to those used in SSRS. In SSIS-RS, children have the response options of not true, little true, a lot true, and very true; caregivers have the response options of never, seldom, often, and almost always. These response options are retained in the SSIS-SEL version.

In addition, the Parent, Teacher, and the Student (13-18-year-old version) Forms ask respondents to rate how important they think each behaviour (asked about in the SSIS items) is, from the response options of not important, important, and critical.[87]

Building on SSIS-RS, the SSIS Social-Emotional Learning Edition (SSIS-SEL) was published in 2017.[80] The SSIS-SEL is designed to measure social-emotional learning, also described as social skills, personal adjustment, social functioning, or social behaviour.[80] The SSIS-SEL was developed to better incorporate elements of social and emotional learning including students managing their emotions, learning to set attainable positive goals, developing strong relationships and effective decision making.[88]

SSIS-SEL assesses students across five different competencies, designed to align with core competencies of the Collaborative for Academic, Social and Emotional Learning. The dimensions are: Self-Awareness, Self-Management, Social Awareness, Relationship Skills, and Responsible Decision Making.

> **Self-Awareness**: Accurate recognition of one's emotions and thoughts and their influence on behavior, including accuracy in the assessment of skills and a well-grounded sense of confidence and optimism.
> **Self-Management**: Effective regulation of one's emotions, thoughts, and behaviors in different situations, including stress management, impulse control, self-motivation, and goal setting.
> **Social Awareness**: Taking the perspective of, and empathizing with, others from diverse backgrounds and cultures, including understanding social and ethical norms for behaviour and recognizing various resources and supports that are available.
> **Relationship Skills**: Establishing and maintaining healthy and rewarding relationships with diverse individuals and groups, including clear communication, active listening, cooperation, resisting peer pressure, conflict negotiation, and seeking/offering help when appropriate.
> **Responsible Decision Making**: Making constructive and respectful choices about personal behaviour and social interactions that are based on consideration of ethical standards, safety concerns, social norms, consequences, and the well-being of self and others.[80 p. 2]

The SSIS-SEL includes Teacher (58 items), Parent (51 items), and Student (46 items) Forms, designed to be completed by the respondent in 10-20 minutes. Each of the three forms can be used to generate scores assessing the five competencies listed above; the Teacher Form additionally generates a score for Academic Competence. These SSIS-SEL forms can be administered on paper or digitally, using a computer, tablet, or handheld device.[80]

Evidence on the validity of the measure for Aboriginal and/or Torres Strait Islander children

The SSRS and SSIS measures were developed and assessed within a United States school population. There are at least two Australian studies that have measured SSIS in a sample that included Aboriginal and/or Torres Strait Islander participants; however, results were not presented specifically for these participants.[89,90] There is evidence of the reliability of the SSIS-SEL for use with other populations, but, to our knowledge, no studies have assessed the validity of the measure for use with Aboriginal and Torres Strait Islander Australians. The Gudaga Goes to School (Gudaga-GtS) Study of Aboriginal children (n = up to 117) collects data on the Academic Competency Scale from the SSIS-SEL (teacher's assessment), but does not include the other competency scales.[91]

Modifications between original measure and measure used in LSIC

LSIC includes a subset of items from the Student and Parent SSIS-RS forms, following the approach used in LSAC. To our knowledge, there is not documentation explaining why this subset of items was chosen for use in LSAC. In both LSIC and LSAC, the main questions are asked, but participants were not asked to rate the importance of the specified behaviours, as occurs in the SSIS-RS.

Five items from the original SSRS Empathy domain were used in LSAC Wave 4 (2010) and Wave 5 (2012), with children in the K Cohort who were aged 10-11 years and 12-13 years, respectively. Children self-reported their answers to questions about how often they exhibit a set of behaviours and feelings. The response options were rarely or never (1), sometimes (2), and very often (3). These items and response options are different to the SSIS-RS items used in subsequent waves.

In Wave 6 (2014), LSAC incorporated items into the child survey (child self-complete) from the revised SSIS-RS, for the K Cohort (aged 14-15 years) only. The corresponding LSIC variable names for the SSIS-RS Empathy Social Skills subscale items are shown in Table 6.1. The response options used in LSIC were: not true (1), little true (2), a lot true (3), and very true (4). The response options are aligned with the intended response options for the Child form, but we note that in the original SSIS-RS, these are coded on a scale from 0-3 rather than 1-4. For the purposes of this Report, the difference in coding does not impact results, because a total SSIS-RS score was not calculated.

In this same wave (LSAC Wave 6), 21 SSIS-RS items were included in the parent or carer survey, for both the B Cohort (aged 10-11 years) and the K Cohort (aged 14-15 years). Parents or carers were asked the complete set of items from the domain of Self-control, and a subset of items from the Cooperation, Assertion, and Responsibility domains. Parents or carers were asked to think about how often the child displayed a set of behaviours over the past two months. Caregivers were asked to rate the frequency of the child's according to the response options of never (1), seldom (2), often (3), and almost always (4). The prompt aligns with the specified time frame on the SSIS-RS Parent form, and the same response options are used.

In Wave 10 of LSIC, SSIS-RS items were included with children and parents or carers, following the approach used in LSAC. LSIC children in the K Cohort only (aged 12-14 years) self-completed the SSIS-RS Student form items (Empathy); parents or carers of children in both cohorts (9-11 years and 12-14 years) completed the Parent form items (complete set of Self-control items; subset of items from the Cooperation, Assertion, and Responsibility domains). Parents or carers were asked *Think about your child in the last two months…* This timeframe is consistent with how the SSIS-RS Parent From is structured, and the same response options were used.

Slight adaptations were made to the SSIS-RS items, in response to pilot testing and feedback from the RAOs and Steering Committee members. The Student form items are presented in Table 6.1 and the Parent form LSIC variable names are presented in Table 6.2. In LSIC, clarifying text is provided, if required, for some of the Parent form SSIS items (8 out of the 21 items), to enable consistency in explanation across participants (Table 6.2). In addition, in the Parent form SSIS items, the clarifier of rarely is provided for the seldom response option.

Table 6.1 and Table 6.2 map the items asked in LSIC onto the SSIS-RS domains and on to the updated SSIS-SEL competencies. The items used in LSIC were based on the SSIS-RS domains, which are organised differently than the SSIS-SEL competencies, so the LSIC items don't neatly align with the SSIS-SEL competencies. The Student form items used in LSIC include 5 of the 7 items from the SSIS-SEL Social Awareness competency, and 1 of the 15 items from the SSIS-SEL Relationships Skills competency. The Parent form items used in LSIC include 10 of the 14 items from the SSIS-SEL Self-Management competency, 1 of the 7 items from the SSIS-SEL Social Awareness competency, 2 of 14 items from the SSIS-SEL Relationship Skills competency, 7 of the 9 items from the SSIS-SEL Responsible Decision Making competency, and 1 item from SSIS-SEL Self-Awareness competency. To our

knowledge, published studies have validated the SSRS and SSIS measures as a whole; we are unaware of any validation studies of the specific subset of items employed in LSIC and LSAC.

Table 6.1. Student form (child self-reported) SSIS items used in LSIC Wave 10, K Cohort

SSIS LSIC variable name child form items used in LSIC SC Survey	SSIS-RS domain	SSIS-SEL competency
csr5_1	Empathy	Relationship Skills
csr5_2	Empathy	Social Awareness
csr5_3	Empathy	Social Awareness
csr5_4	Empathy	Social Awareness
csr5_5	Empathy	Social Awareness
csr5_6	Empathy	Social Awareness

Table 6.2. Parent form (parent- or carer-reported) SSIS items in LSIC Wave 10 (B and K Cohorts)

SSIS LSIC variable name parent form items used in LSIC P1 Survey	SSIS-RS domain	SSIS-SEL competency
asr1_a	Self-Control	Self-Management
asr1_b	Self-Control	Self-Management
asr1_c	Self-Control	Self-Management
asr1_d	Self-Control	Relationship Skills
asr1_e	Self-Control	Self-Management
asr1_f	Self-Control	Self-Management
asr2_g	Self-Control	Self-Management
asr2_a	Cooperation	Responsible Decision Making
asr2_b	Cooperation	Self-Management
asr2_c	Cooperation	Relationship Skills
asr2_d	Cooperation	Self-Management
asr3_a	Assertion	Self-Management
asr3_b	Assertion	Self-Management
asr3_c	Assertion	Social Awareness
asr3_d	Assertion	Self-Awareness
asr3_e	Assertion	Responsible Decision Making
asr4_a	Responsibility	Responsible Decision Making
asr4_b	Responsibility	Responsible Decision Making
asr4_c	Responsibility	Responsible Decision Making
asr4_d	Responsibility	Responsible Decision Making
asr4_e	Responsibility	Responsible Decision Making

Coding and scoring

The SSIS is designed to be completed by multiple informants, including the student, parent(s) or carer(s), and teacher(s) at the same point in time. The SSIS manual describes how assessments from multiple informants can be evaluated. The scores for each form are generated by summing the scores of each of the relevant items; the manual provides information for how missing values can be handled when the full set of items are used. When the whole set of items is used, an online scoring system can generate raw scores and percentile ranks according to the normative data from the United States.

The SSIS-RS manual provides information about how to standardise scores to enable comparability across scales. A similar approach can be used for the SSIS-SEL forms.[87] Norms have been created based on a national sample of 4,700 children aged 3-18 years in the United States from 2006-2007. Norms have been created for three age groups (3-5, 5-12, and 13-18 years), overall and by gender. In general, the SSIS-RS manual, recommends the use of gender-specific norms given gender differences in social-emotional learning development.[87] The derived standardised scores have a mean of 100 and a Standard Deviation (SD) of 15, with a higher score indicating a higher level of social-emotional learning. Scores within 1SD of the mean are considered 'Average' (score: 85 - 115); scores > 1-2SDs below or above the mean are considered 'Below/Above Average' (score: 70 - 84, 116 - 130, respectively); scores > 2SDs below or above the mean are considered 'Well-below/Above Average'. A standard Social Skills score below 85 suggests there is a need for social skills training or behavioural intervention.[87]

We lack evidence on whether these norms and associated judgments are appropriate for use with Aboriginal and Torres Strait Islander children. However, in any case, this advice is not applicable to LSIC given the limited number of items included in LSIC. The manual does not include any information about generating scores on partial subscales or how to interpret scores from individual subscales, when not included as part of the whole form.

Results: assessment of SSIS in LSIC

Analysis was conducted separately for each cohort. The results varied slightly between the cohorts so results are reported separately, and a combined cohort analysis was not conducted.

A PCA rather than CFA was conducted, given that the items used in LSIC were an abridged version of the original SSIS-RS, and that the items have not been previously validated with an Aboriginal and Torres Strait Islander sample.

Acceptability

Overall, there was a moderate extent of non-response. For both parents or carers and children there was a higher rate of non-responses coded as don't know compared to refuse. Overall, children had 26.5 item non-responses per 100 surveys and parents or carers had 35.4 item non-responses per 100 surveys.

> **Key points**: Overall, the extent of non-response to SSIS items was moderate in this sample, indicating that there may be some concerns about interpretability and/or acceptability of some items.

Construct validity

Suitability of the data for a PCA was assessed. For the child self-reported Empathy items, all inter-item correlations were ≥ 0.3 for both cohorts and retained for the PCA. For the P1-reported SSIS-RS items, most inter-item correlations were ≥ 0.3. For the B Cohort, most P1-reported SSIS-RS items had an inter-item correlation ≥ 0.3. For the K Cohort, the P1-reported SSIS-RS items had a low number of inter-item correlations ≥ 0.3 within the Assertion subscale. Specifically two items from the assertion subscale (asr3_a and asr3_b) not correlate above the 0.3 threshold with two other items from the assertion subscale (asr3_d and asr3_e).

For both cohorts, the KMO measure of sampling adequacy was above the recommended 0.6 (KMO ≥ 0.88 for children and KMO ≥ 0.90 for parents or carers) and the Bartlett's test of sphericity was significant for parents or carers and children. This indicated that the data were suitable for a PCA; all items were retained. A PCA was conducted on the SSIS-RS items, by informant and cohort.

Child self-reported items (Empathy)

Based on the parallel analysis, screeplot, and Kaiser's criterion, only one factor was retained for the child self-reported (Empathy) items, as was hypothesised. The PCA (orthogonal rotation) for the child self-reported items confirmed that all items loaded onto one component > 0.3 (results not shown).

Parent- or carer-reported items: B Cohort

Three- component and four-component solutions were retained for parents or carers of children in the B Cohort (Table 6.3). The three-component solution for the B Cohort explained 55.3% of the variance. Three Cooperation items did not load onto a component (asr2_a, asr2_b and asr2_c) (results not shown). The four-component solution explained 60.8% of the variance. The loading of items on the four components aligned with the current SSIS-RS subscale structure, except for one Self-control item (asr1_a) which did not load onto any factor > 0.3. The four-component structure supported the hypothesised scales and was retained (Table 6.3).

Table 6.3. PCA for the SSIS-RS measure, B Cohort Wave 10 (four-component solution)

	Component			
	1	**2**	**3**	**4**
Self-control				
asr1_a				
asr1_a	0.397			
asr1_a	0.317			
asr1_a	0.351			
asr1_a	0.438			
asr1_a	0.459			
asr1_a	0.333			
Cooperation				
asr2_a		0.520		
asr2_b		0.517		
asr2_c		0.337		
asr2_d		0.466		
Assertion				
asr3_a				0.398
asr3_b				0.439
asr3_c				0.482
asr3_d				0.386
asr3_e				0.484
Responsibility				
asr4_a			0.505	
asr4_b			0.518	
asr4_c			0.342	
asr4_d			0.381	
asr4_e			0.339	

Parent- or carer-reported items: K Cohort

Three- to five-component solutions were retained for parents or carers of children in the K Cohort. The three-component solution for the K Cohort explained 56.7% of the variance (results not shown). Five items did not load onto any components: one self-control (asr1_a), and four responsibility (asr4_a, asr4_c, asr4_d, and asr4_e)

The four-component solution for the K Cohort explained 62.5% of the variance (Table 6.4). Four items did not load onto any components: one self-control (asr1_a), one assertion (asr3_d) and two responsibility (asr4_d, and asr4_e). The Self-control and the Cooperation subscales emerged as two distinct components; however, results were not as expected for the assertion and responsibility factors. Two of the Assertion items (asr3_a, and asr3_b), emerged as a two-item component. Two other Assertion items (asr3_c, and asr3_e), loaded on to the Responsibility factor.

The five-component solution explained 67.6% of the variance (Table 6.5). Three items did not load onto any components: one self-control (asr1_a) and two responsibility (asr4_d, and asr4_e); these three items also did not load in the four-component solution. In the five-component solution, the Self-control, Cooperation, and Responsibility subscales emerged as three distinct components. The Assertion items emerged as two distinct components, the two-item component that arose in the four-component solution, and a three-item component (see Table 6.5).

Table 6.4. PCA for the SSIS-RS measure, K Cohort Wave 10 (four-component solution)

	Component			
	1	**2**	**3**	**4**
Self-control				
asr1_a				
asr1_a		0.427		
asr1_a		0.373		
asr1_a		0.312		
asr1_a		0.429		
asr1_a		0.403		
asr1_a		0.385		
Cooperation				
asr2_a			0.443	
asr2_b			0.488	
asr2_c			0.360	
asr2_d			0.412	
Assertion				
asr3_a				0.628
asr3_b				0.590
asr3_c	0.386			
asr3_d				
asr3_e	0.306			
Responsibility				
asr4_a	0.441			
asr4_b	0.448			
asr4_c	0.358			
asr4_d				
asr4_e				

Items with a loading < 0.3 were omitted from the table.

The two SSIS-RS Assertion items that emerged as a two-item component (in both the four-component solution and the five-component solution) are both grouped under the Self-Management competency in the updated SSIS-SEL scheme. The remaining items map on to three different competencies: social awareness (asr3_e), self-awareness (asr3_d), and responsible decision making (asr3_e). It is difficult to further assess the fit of the data against the SSIS-SEL competencies given the items included in LSIC.

> ***Key points***: The child self-reported Empathy items performed as expected, emerging as a single construct. For the B Cohort, the parent- or carer-reported SSIS-RS items performed as expected, supporting the hypothesised scale structure; only one item (Self-control) did not perform well. For the K Cohort, results were less clear; three to four items did not load onto any components, and the Assertion subscale was particularly problematic.

Table 6.5. PCA for the SSIS-RS measure, K Cohort Wave 10 (five-component solution)

	Component				
	1	**2**	**3**	**4**	**5**
Self-control					
asr1_a					
asr1_a	0.425				
asr1_a	0.349				
asr1_a	0.312				
asr1_a	0.441				
asr1_a	0.430				
asr1_a	0.374				
Cooperation					
asr2_a		0.455			
asr2_b		0.537			
asr2_c		0.386			
asr2_d		0.469			
Assertion					
asr3_a					0.678
asr3_b					0.655
asr3_c				0.443	
asr3_d				0.580	
asr3_e				0.610	
Responsibility					
asr4_a			0.568		
asr4_b			0.552		
asr4_c			0.419		
asr4_d					
asr4_e					

Items with a loading < 0.3 were omitted from the table.

Internal consistency

The internal consistency for the child self-rated Empathy scale had an acceptable alpha score for each cohort (B Cohort: α = 0.84; K Cohort: α = 0.90).

Internal consistency was calculated for the original hypothesised SSIS-RS subscales (Table 6.6). Internal consistency was acceptable for all subscales for both cohorts.

Table 6.6. Cronbach's alpha of the parent- or carer-reported SSIS subscales, by cohort

	α			
Cohort	**Self-control**	**Cooperation**	**Assertion**	**Responsibility**
B Cohort	0.85 ✓	0.86 ✓	0.75 ✓	0.86 ✓
K Cohort	0.84 ✓	0.88 ✓	0.74 ✓	0.89 ✓

X = unacceptable (alpha < 0.6). ~ = approaching acceptable (alpha 0.6 - < 0.7). ✓ = acceptable (alpha 0.7 - 0.9).

In addition, Cronbach's alpha was calculated according to the four-component solution (B Cohort) and the five-component solution (K Cohort), based on the PCA results (Table 6.7). Items with a low loading (< 0.3) were excluded from this analysis. In addition, Cronbach's alpha was calculated for the two-item component (Component 5) that arose in the five-component solution for the K Cohort. The internal consistency remained acceptable (0.7 - 0.9), except for the Assertion subscale (K Cohort) which was

reduced to approaching acceptable (0.69). Internal consistency was acceptable for the two-item component (Component 5).

Table 6.7. Cronbach's alpha of the parent or carer- reported SSIS subscales, adjusted based on PCA results, by cohort

Cohort	α				
	Self-control	Cooperation	Assertion	Responsibility	Component 5 (two-item)
B Cohort	0.83 ✓	0.86 ✓	0.75 ✓	0.86 ✓	N/A
K Cohort	0.82 ✓	0.88 ✓	0.69 ~	0.85 ✓	0.81* ✓

r = 0.64. X = unacceptable (alpha < 0.6). ~ = approaching acceptable (alpha 0.6 - < 0.7). ✓ = acceptable (alpha 0.7 - 0.9).

Key points: For both cohorts, one item (asr1_a) did not load as expected onto the Self-control subscale, but the scale internal consistency was not improved by excluding the item. For the K Cohort, multiple items did not load onto any factors; however, the internal consistency of subscales was not improved by excluding these items. This suggests that these items are not negatively influencing the internal consistency of the subscales.

Recommendations

1. **Evidence of validity:** Results were broadly consistent with the intended SSIS-RS structure. The child-reported Empathy subscale performed well for the K Cohort (these items were not asked with the B Cohort). The Cooperation and Self-control subscales appeared to be (psychometrically) valid, reliable measures for both cohorts. The Responsibility and Assertion subscales performed well for the B Cohort, but less well for the K Cohort. We recommend analysing data from the two cohorts separately when using these two subscales.
   1.1. However, we are unaware of any qualitative assessment of the meaning and relevance of the SSIS-RS items within the Aboriginal and Torres Strait Islander context. Further qualitative item validation research across contexts and language groups is required to accept the validity of the measure for use with Aboriginal and Torres Strait Islander children.[19,21]
   1.2. Accordingly, SSIS-RS data with this population must be interpreted with caution, and should be interpreted within context.
   1.3. Researchers using the SSIS-RS need to be explicit about the limitations of using this measure with Aboriginal and Torres Strait Islander peoples.
   1.4. It is not appropriate to compare the SSIS-RS between Aboriginal and Torres Strait Islander and non-Indigenous children; we cannot assume that the SSIS-RS is conceptually equivalent (measuring the same thing) in these populations.

2. **Caution in interpreting the Assertion subscale for the K Cohort:** The Assertion subscale performed poorly for the K Cohort, and results should be interpreted with caution.

7. Short Infant Temperament Questionnaire (SITQ)

Background

Purpose of SITQ

There is evidence from other populations that temperament may be important to children's mental health and wellbeing;[92,93] however, there is limited research on the concept of temperament within the Aboriginal and Torres Strait Islander population.[93] Temperament has been defined as 'stable, constitutionally based characteristics or behavioural styles that may be evident from birth (see review by Sanson et al.[94]). While they have biological roots, they are not set in stone; environments and contexts may promote or discourage the display of particular traits'.[93 p. 93] Research has identified three broad elements of child temperament: Approach-sociability (Approach), the extent to which children are comfortable engaging in new situations or with new people; Persistence, the child's ability to regulate his/herself and to complete tasks; and Reactivity, the child's level of intensity and emotional instability.[93] Temperament is generally assessed through parent- or carer-report on written questionnaires.[93] The Short Infant Temperament Questionnaire (SITQ) is commonly used internationally as measure of temperament for children aged 3-8 years.[95]

The SITQ includes 30 items intended to measure different dimensions of temperament: Approach (7 items), Rhythmicity (7 items), Inflexibility (9 items), and Persistence (7 items). Parents or carers reported how frequently they observed the child exhibiting the behaviour described in each item, from the response options of: almost never (1), not often (2), usually does not (3), usually does (4), frequently (5), and almost always (6).

Modifications between original scale and scale used in LSIC

A reduced set of SITQ items are used in LSIC (Table 7.1); this subset of items is also used in LSAC. In addition, the administration of SITQ is modified for use LSIC. In LSIC, the SITQ is administered orally by the interviewer, rather than the parent or carer self-completing a written questionnaire.[93] To reflect this change, the SITQ statements were rephrased as questions.

The reduced set of SITQ items used in LSIC comprises three sub-scales (Approach, Persistence, and Reactivity) with four items each. None of the subscales include the full set of items from the SITQ (missing 3-5 items for each subscale).

Table 7.1 Parent form (parent- or carer-reported) SITQ items in LSIC Wave 10 (B and K Cohorts)

SITQ Parent form items used in LSIC P1 Survey	SITQ Sub-scale
1. (He/she) is shy with strange adults? *(reverse coded)*	Approach
2. When (he/she) starts a project such as a puzzle or a model, (he/she) works on it without stopping until it is completed even if it takes a long time?	Persistence
3. If (he/she) wants a toy or sweet while shopping, (he/she) will easily accept something else instead?	Reactivity
4. (He/she) is shy when first meeting new children? *(Reverse Coded)*	Approach
5. (He/she) likes to complete one task or activity before going onto the next?	Persistence
6. When (he/she) is angry about something, it is difficult to sidetrack (him/her)?	Reactivity
7. When in a park or visiting, (he/she) will go up to strange children and join in their play	Approach
8. (He/she) stays with an activity (e.g. puzzle, construction kit, reading) for a long time?	Persistence
9. When shopping together, if I do not buy what (he/she) wants (e.g. sweets, clothing) (he/she) cries and yells?	Reactivity
10. When unknown adults visit our home, (he/she) is immediately friendly and approaches them?	Approach
11. If (he/she) is upset, it is hard to comfort him/her?	Reactivity
12. When a toy or game is difficult, (he/she) quickly turns to another activity? *(reverse coded)*	Persistence

Individual item scores and derived sub-scale (dimension) scores are provided in LSIC. Sub-scale scores are created by reverse coding some items (as described above) and then averaging all items within the dimension (range: 1-6). This generates an 'easy-difficult' score, with a higher score indicating a more 'difficult' temperament.[96] Sub-scale scores are not derived where three or four of the items from the subscale are missing.[97]

Evidence on the validity of the measure for Aboriginal and/or Torres Strait Islander children

Little et al. used data from LSIC to examine 'the structure of temperament in Indigenous children as well as how temperament, along with parenting style, might be linked to their later emotional and behavioural adjustment'.[93 p. 92] The authors analysed data from the Longitudinal Study of Australian Children (LSAC) for comparison, to explore whether similar elements are relevant to temperament in Aboriginal and Torres Strait Islander children and non-Indigenous children, and to compare the prevalence of temperament-related characteristics among the Aboriginal and Torres Strait Islander children and non-Indigenous children.[93]

Based on research from other populations, the authors expected that the distribution of responses to each item would roughly follow a normal distribution, with the majority of responses of the middle categories usually does not (3), usually does (4), and fewer responses at the extremes almost never (1), almost always (6).[93] However, they found a more uniform distribution of responses across categories in LSIC; for many items, there was a trimodal distribution with peaks at almost never (1), usually does (4), and almost always (6). Based on this response pattern, the authors considered that it would be inappropriate to form dimension sub-scale scores by averaging responses across the items.[93] While it is important to further investigate the distribution of responses, we note that the existence of normally distributed items is not a requirement for PCA.[34]

The authors propose several potential explanations for the difference in response patterns; one key hypothesis is that the interviewer administration of the measure in LSIC (rather than written P1 self-complete) might contribute to the difference in response patterns between the LSIC and the LSAC sample.[93] For example, respondents may not have answered the questions specifically using the intended response options as follows: almost never, not often, usually does not, usually does, frequently, and almost always. As such, interviewers may have had to interpret the P1's response and assign it to what they perceived to be the most relevant of the available options. This hypothesis is supported by our focus group with the LSIC RAOs, who described needing to "interpret" the answers that respondents provided. It may have been difficult for respondents and/or interviewers to differentiate between the response options of almost never, not often, and usually does not, and between the response options of frequently and almost always; if the former responses were most often recorded as almost never and the latter as almost always, this could result in the observed trimodal distribution.

Little et al. propose transforming the 6-category response scale to a 3-category scale: not much (1), combining almost never and not often; sometimes (2), combining usually does not and usually does, and often (3), combining frequently and almost always.[93] Use of this collapsed categorisation to derive each dimension sub-scale score generated a relatively more normal distribution (range:1-3, with 1 representing a low level of that trait, and 3 representing a high level of that trait), though the authors note the distribution was relatively flat.[93]

Within the LSIC cohort, Little et al. identified significant associations between temperament sub-scales and risk of Emotional, Hyperactivity, and Conduct Problems, as measured by the SDQ.[93] Children with higher Approach scores (more outgoing) were less likely to have emotional problems than children with lower Approach scores (more shy). Children with higher compared to lower Persistence scores

had a significantly lower risk of inattention/hyperactivity problems and conduct problems; no association was observed with emotional problems.

Results: assessment of SITQ in LSIC

The SITQ subscales were administered to the B Cohort at Wave 5 and K Cohort at Wave 2 when children were aged 4 to 6 years. As children were of equal ages when the instrument was administered, the current results present each cohort separately and then the cohorts combined.

Acceptability

Overall there was a low extent of non-response, with an average of 9.5 item non-responses per 100 surveys. This likely suggests that overall, respondents were satisfied with the items and willing to provide a response.

> **Key points**: Overall, the extent of non-response to SITQ items was low in this sample, indicating likely acceptability of the items.

Response distribution

We explored the distribution of responses across SITQ items in the B Cohort (age 4-6 years, Wave 5) to the K Cohort (age 4-6 years, Wave 2). We generally observed a similar pattern in the B Cohort as Little et al. had observed in the K Cohort, with three peaks in responses, at almost never (1), usually does (4), and almost always (6). As an example, the distribution of responses for three items, one from each subscale, are presented in Figure 7.1. We note that when we generate a total score for each subscale, the distribution is relatively normal within the B Cohort, and less so in the K Cohort (Figure 7.2). We have not reduced the SITQ items to a 3-point scale; we have retained the 6-category scoring for all analyses.

> **Key points**: Response options of almost never (1), usually does (4), and almost always (6) were most common, for both the B Cohort and the K Cohort. However, when subscale scores are calculated, the distribution becomes more normal, particularly for the B Cohort.

Figure 7.1. Distribution of response options to three SITQ items among children in the B Cohort (Wave 5) and in the K Cohort (Wave 2)

Temperament- SC likes to complete one task before going on to the next

Temperament- SC wants a sweet, but will easily accept a substitute
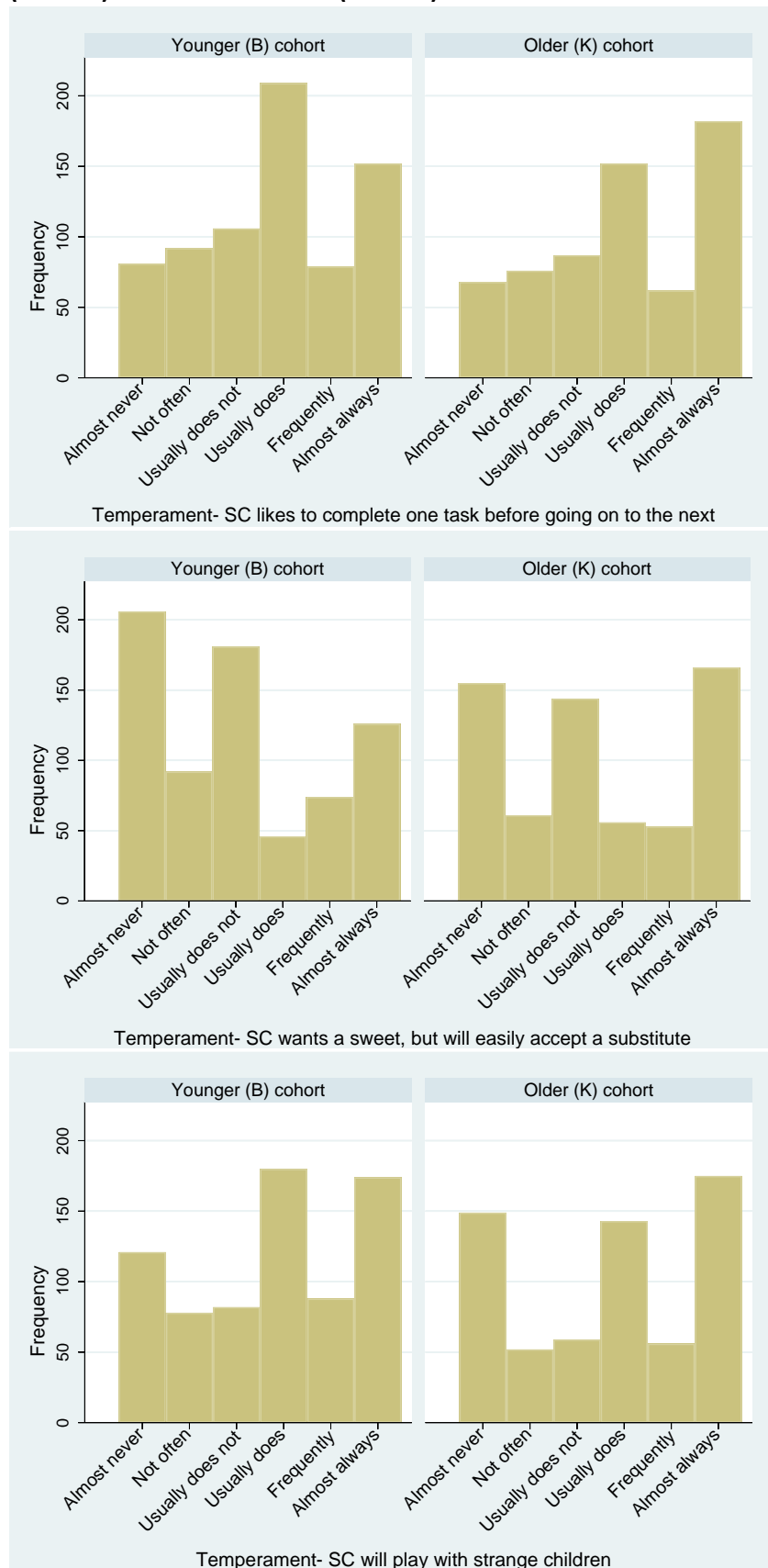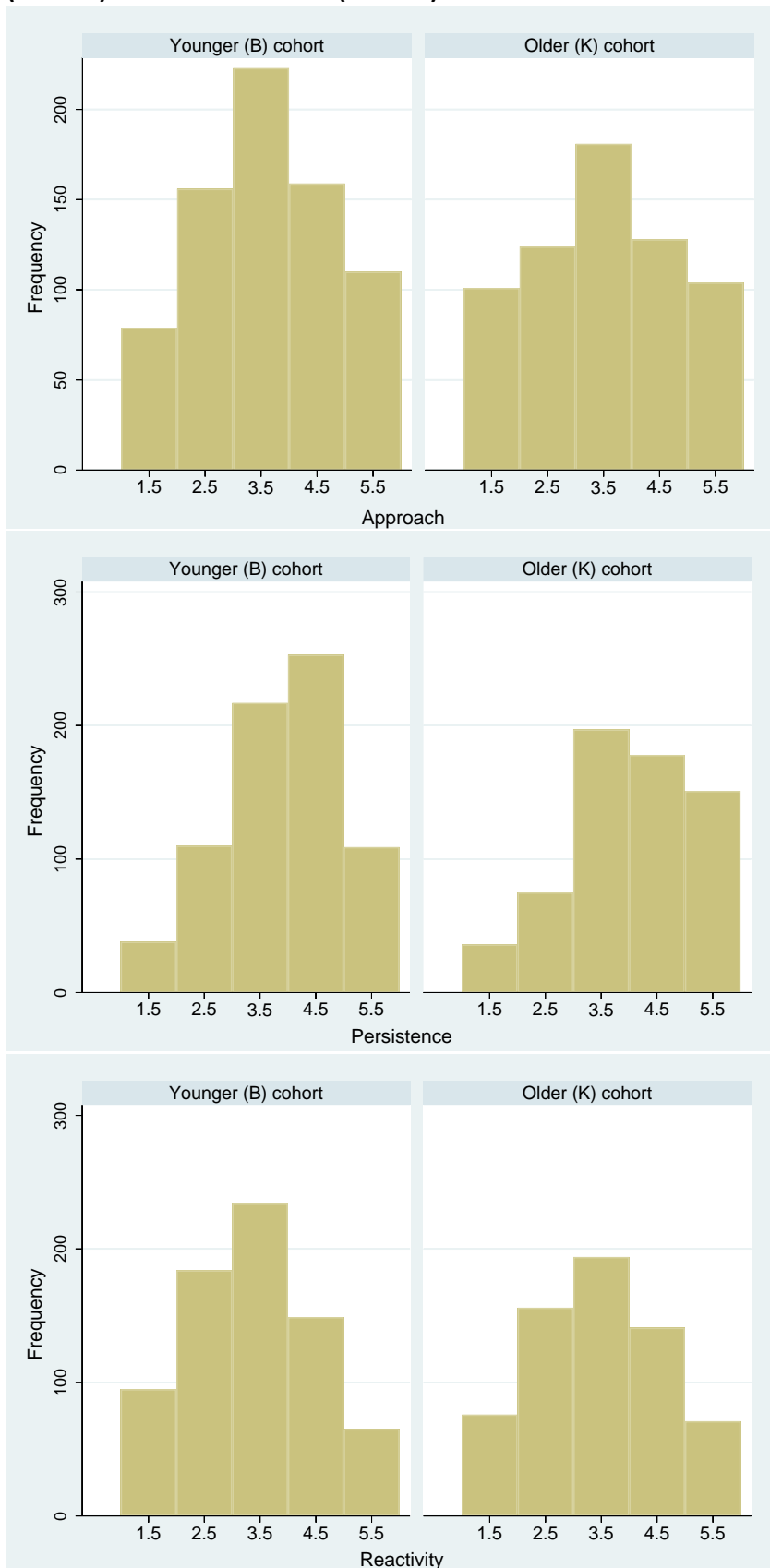
Temperament- SC will play with strange children

Figure 7.2. Distribution of total Approach, Persistence, and Reactivity subscale scores in the B Cohort (Wave 5) and in the K Cohort (Wave 2)

Construct validity

A PCA was employed to explore the number of reliable components from the items, and to examine the relationship between the items and the components. This exploratory approach was employed as the current subscales were not the complete set of items and the intended three-factor model could not be assumed.

The existence of normally distributed items is not a requirement for PCA;[34] therefore, the non-normal distribution of some SITQ items in LSIC does not preclude PCA.

All items correlated ($r \geq 0.3$) with at least one other item, with the exception of the item, *When a toy or game is difficult, (he/she) quickly turns to another activity*? (Persistence). This item did not correlate above the 0.3 threshold with any other item, for both cohorts. This item was consequently excluded from analyses. For both cohorts the Bartlett's test of sphericity indicated the correlations were not too low for a PCA (p-value < 0.001) and the KMO Measure of Sampling Adequacy indicated that data were suitable for PCA (0.68 for each cohort).

A PCA was conducted on the remaining 11 items (excluding *when a toy or game is difficult, (he/she) quickly turns to another activity*?). Based on the parallel analysis, screeplot, and Kaiser's criterion, three- and four-component solutions were retained and analysed for both cohorts.

B Cohort

The three-component solution for the B Cohort (n = 701) explained 53.9% of the variance and the loading of items on the three components aligned with the hypothesised SITQ subscale structure. The four-component solution explained 62.3% of the variance. The Approach and Persistence subscales appeared as distinct components; however items from the Reactivity subscale loaded onto two distinct components (B Cohort results not shown).

K Cohort

A similar result was found for the K Cohort (n = 607). The three-component solution explained 54.5% of the variance. Similar to the B Cohort, the loading of items on the three factors aligned with the hypothesised SITQ subscale structure. The four-component solution for the K Cohort explained 63.6% of the variance. The Approach and Persistence subscales appeared as distinct components; however items from the Reactivity subscale loaded onto two distinct components, as occurred with the B Cohort (K cohort results not shown).

Both cohorts combined

Given the similar pattern of results for the B Cohort and the K Cohort, analysis was repeated for both cohorts combined (n = 1308). The results mirrored those reported above. The three-component solution explained 54.0% of the variance, and the loading of items on the three components aligned with the hypothesised SITQ subscale structure. See Table 7.2. The Approach items loaded above the 0.3 threshold on Component 1, the Persistence items loaded on Component 2 and the Reactivity items loaded onto Component 3.

Table 7.2. PCA for parent or carer reported SITQ, B and K Cohorts combined aged 4.5-6 years (three-component solution)

	Component		
	1	2	3
(He/she) is shy with strange adults?	0.522		
(He/she) is shy when first meeting new children?	0.507		
When in a park or visiting, (he/she) will go up to strange children and join in their play?	0.473		
When unknown adults visit our home, (he/she) is immediately friendly and approaches them?	0.496		
When (he/she) starts a project such as a puzzle or a model, (he/she) works on it without stopping until it is completed even if it takes a long time?		0.555	
(He/she) likes to complete one task or activity before going onto the next?		0.570	
(He/she) stays with an activity (e.g. puzzle, construction kit, reading) for a long time?		0.557	
If (he/she) wants a toy or sweet while shopping, (he/she) will easily accept something else instead?			0.457
When (he/she) is angry about something, it is difficult to side track (him/her)?			0.507
When shopping together, if I do not buy what (he/she) wants (e.g. sweets, clothing) (he/she) cries and yells?			0.559
If (he/she) is upset, it is hard to comfort him/her?			0.460

The four-component solution explained 62.6% of the variance (Table 7.3). The loading patterns were similar to those observed within the B Cohort and K Cohort. The Approach (Component 1) and Persistence (Component 2) subscales appeared as distinct components; however, items from the Reactivity subscale loaded onto two separate components. Items *If (he/she) wants a toy or sweet while shopping, (he/she) will easily accept something else instead?* and, *When shopping together, if I do not buy what (he/she) wants (e.g. sweets, clothing) (he/she) cries and yells?,* loaded onto Component 3, while *When (he/she) is angry about something, it is difficult to sidetrack (him/her)?* and *If (he/she) is upset, it is hard to comfort him/her?*, loaded onto Component 4. Our PCA findings with 6-category items were similar to those of Little et al., who reduced the 6-point scale to a 3-point scale.

Table 7.3. PCA for parent or carer reported SITQ, B and K Cohorts combined aged 4.5-6 years (four-component solution)

	Component			
	1	2	3	4
(He/she) is shy with strange adults?	0.518			
(He/she) is shy when first meeting new children?	0.503			
When in a park or visiting, (he/she) will go up to strange children and join in their play?	0.476			
When unknown adults visit our home, (he/she) is immediately friendly and approaches them?	0.502			
When (he/she) starts a project such as a puzzle or a model, (he/she) works on it without stopping until it is completed even if it takes a long time?		0.581		
(He/she) likes to complete one task or activity before going onto the next?		0.575		
(He/she) stays with an activity (e.g. puzzle, construction kit, reading) for a long time?		0.561		
If (he/she) wants a toy or sweet while shopping, (he/she) will easily accept something else instead?			0.749	
When shopping together, if I do not buy what (he/she) wants (e.g. sweets, clothing) (he/she) cries and yells?			0.630	
When (he/she) is angry about something, it is difficult to side track (him/her)?				0.659
If (he/she) is upset, it is hard to comfort him/her?				0.671

> ***Key points***: The three-component solution was consistent with the hypothesised SITQ subscale structure; however, the four-component solution highlighted some potential issues with the Reactivity scale. Results were similar for the B Cohort, K Cohort, and the B and K Cohorts combined.

Internal consistency

Internal consistency was acceptable, or approaching acceptable, for the Approach and Persistence subscales, and unacceptably low for the Reactivity subscale (Table 5.1). This was expected as the Reactivity scale appeared to perform better as two distinct subscales (as described above).

Table 5.1. Internal consistency of SITQ subscales, by Cohort and overall

Cohort	α		
	Approach	Persistence (3 items)	Reactivity
B Cohort	0.70 ✓	0.71 ✓	0.60 X
K Cohort	0.69 ~	0.77 ✓	0.55 X
Both Cohorts	0.69 ~	0.74 ✓	0.57 X

X = unacceptable (alpha < 0.6). ~ = approaching acceptable (alpha 0. 6 - < 0.7). ✓ = acceptable (alpha 0.7 - 0.9).

> **Key points**: Internal consistency was broadly acceptable for the Approach subscale and the reduced Persistence subscale, but not acceptable for the Reactivity subscale.

Recommendations

1. **Evidence of validity:** The SITQ Approach and Persistence subscales appear to be (psychometrically) valid and reliable measures of a child's temperament. These are also strengths-based measures.
   1.1. However, we are unaware of any qualitative assessment of the meaning and relevance of the SATI items within the Aboriginal and Torres Strait Islander context. Further qualitative item validation research across contexts and language groups is required to accept the validity of the measure for use with Aboriginal and Torres Strait Islander children.[19,21]
   1.2. Accordingly, SATI data with this population must be interpreted with caution, and should be interpreted within context.
   1.3. Researchers using the SATI need to be explicit about the limitations of using this measure with Aboriginal and Torres Strait Islander peoples.
   1.4. It is not appropriate to compare the SITQ between Aboriginal and Torres Strait Islander and non-Indigenous children; we cannot assume that the SITQ is conceptually equivalent (measuring the same thing) in these populations.

2. **Exercise caution in using the Reactivity subscale:** We recommend caution in using the Reactivity subscale.
   2.1. When this subscale is used, we recommend data users remove the item, *When a toy or game is difficult, (he/she) quickly turns to another activity?*. The resulting subscale score and interpretation will remain unchanged because the subscale score is an average (rather than the sum) of all items within the subscale.

3. **Combining cohorts:** These results suggest that it is appropriate for data users to analyse data from the cohorts individually, or to combine the two cohorts to create a pooled sample of children of the same age range.

4. **Maintaining the 6-point response scale:** Our findings indicate that it is appropriate to retain the original 6-point response scale when using total subscale scores; rescaling to a 3-point response scale is not required.

8. School-Age Temperament Inventory (SATI)

Background

Purpose of SATI

Temperament is defined as 'individual differences in attentional, emotional, and behavioural self-regulation, along with the relative level of emotional reactivity, which together give a unique flavour to an individual'.[98 p.3] Understanding temperament can help in understanding how children react to different life experiences.[99] Differences in temperament, along with their environment, may contribute to children's development.[93,94,98] A child's temperament generally remains stable across the life course; however, temperament can be modified and adapted through experiences.[98]

A child's temperament is generally assessed by parents or carers, as they have the greatest exposure to the child's behaviours across different environments,[94,99] though parent or carer perceptions of temperament can be biased by their expectations of their child's behaviour and personality.[99] Many temperament measure have been developed with a relatively homogenous samples from the United States.[99]

The 38-item School Age Temperament Inventory (SATI) was developed by McClowry in 1995 to assess temperament among children aged 8-11 years.[100] The SATI is designed to capture four broad dimensions of temperament:
1. **Negative Reactivity**: the intensity and frequency of expression of negative affect;
2. **Task Persistence**: the degree of self-direction a child exhibits in completing tasks;
3. **Approach/Withdrawal**: the child's initial response to new people and situations; and,
4. **Activity**: the child's extent of large motor activity.[99]

The SATI was developed to help understand how different temperaments might be risk or protective factors in psychological development for school aged children.[100]

Evidence on the validity of the measure for Aboriginal and/or Torres Strait Islander children

The SATI has been highlighted as a reliable and valid tool for assessing children's temperament in other populations.[99,100] McClowry has acknowledge that further validation work is required to determine the measure's validity with culturally diverse populations.[100] There is some evidence to indicate that this tool is adequate for the use within the total Australian population.[98,99]

To our knowledge, there is no evidence of the validity of the measure for use with Aboriginal and/or Torres Strait Islander children, and there has been limited validation work conducted with populations from culturally diverse backgrounds.

Modifications between original measure and measure used in LSIC

The SATI was used in the Australian Institute of Family Studies Australian Temperament Project 1983-2000.[98] Before use with this Australian population, adaptations and slight wording modifications were made to the SATI. For example, *Runs to get where he/she wants to go* was changed to *Moves fast to get where s/he want to go*, and *Runs when entering or leaving the house* was changed to *Prefers active outdoor activities to quiet indoor ones*.[99 p.179] To our knowledge, the Australian Temperament Project have not published a full list of the SATI items used in the project or details on the processes of modifying the instrument. A subset of items from the SATI have been used in LSAC, retaining the modifications from the Australian Temperament Project. Items were used with the K Cohort in LSAC Waves 3-5 (aged 8-15 years), and with the B Cohort in LSAC Waves 5 and 6 (aged 8-11 years). The same subset of SATI items used in LSAC was also used in LSIC (Table 8.1). Given that only a subset of items from SATI are used in LSIC, these items cannot be scored according to standard procedures for the full SATI measure.

The SATI items were used when the Study Children were aged 7-9 years: in Wave 5 for the K Cohort, and in Wave 8 for the B Cohort. LSIC includes 12 items total, including items from three of the four domains: approach/withdrawal, task persistence, and negative reactivity. The parent or carer was prompted, *Please tell me which answer best describes how often <Study Child>'s behaviour matches the behaviour described in each question ....* The response options were: never (1), rarely (2), [about] half the time (3), frequently (4), and always (5).

Table 8.1. SATI items included in LSIC Wave 5 and Wave 8, original SATI item wording, and corresponding SATI temperament dimension

LSIC item	SATI item wording	Temperament dimension
1. Cries or complains loudly to a disappointment or failure (e.g. if <he/she> is not able to do something)	Reacts strongly to a disappointment.	Negative Reactivity
2. Needs to be reminded to do homework *(Reverse scored)*	Remembers to do homework.	Task Persistence
3. Is shy with adults <he/she> doesn't know *(Reverse scored)*	Is shy with adults he/she doesn't know.	Approach/Withdrawal
4. When angry, yells or snaps at others	When angry, yells or snaps at others.	Negative Reactivity
5. Goes back to the task at hand (chore, housework, etc.) after an interruption	Goes back to the task at hand.	Task Persistence
6. When meeting new children, acts shy or uncomfortable *(Reverse scored)*	Bashful when meeting new children.	Approach/Withdrawal
7. Sulks when <he/she>'s behaved badly (done the wrong thing)	Moody when corrected for misbehavior.	Negative Reactivity
8. Does homework without anyone reminding <him/her>	Remembers to do homework.	Task Persistence
9. Seems uncomfortable when at someone's house for the first time *(Reverse scored)*	Seems uncomfortable at someone's house.	Approach/Withdrawal
10. Gets very upset if anyone tells <him/her> off	Gets mad when mildly criticized.	Negative Reactivity
11. Has difficulty completing assignments (homework, chores etc.) *(Reverse scored)*	Has difficulty completing assignments.	Task Persistence
12. Approaches children <his/her> age even when <he/she> doesn't know them	Approaches children his/her age.	Approach/Withdrawal

Results: assessment of SATI in LSIC

Acceptability

Overall, there was an average of 24.3 item non-responses per 100 surveys. Three items from the Persistence subscale had the greatest non-response rate. Two Persistence items *Needs to be reminded to do homework or chores*, and *Does homework or chores without anyone reminding <him/her>* had an average of 6.2 item non-responses per 100 surveys; a third Persistence item *Has difficulty completing assignments (homework, chores etc.) finishing things* averaged 4.5 item non-responses per 100 surveys. A cross-tabulation of a Child Care and Early Education item (*P1 checks or helps SC with homework*) and *Does homework or chores without anyone reminding <him/her>* revealed approximately 15% of parents or carers of children in the K-cohort and 20% of the B-cohort reported the child did not receive homework. This may have influenced the current results.

For all three of these items, responses of 'refused' were more common than responses of don't know for parents or carers of children in the B Cohort, but the reverse was true for parents or carers of children in the K Cohort.

Key points: Overall, the extent of non-response to SATI items was moderate in this sample, indicating that there may be some concerns about interpretability and/or acceptability for some items. The extent of non-response was particularly high for three of the Persistence items.

Construct validity

For both cohorts, the correlation between items from different subscales was low. However, within each subscale, all inter-item correlations were above the 0.3 threshold. The Bartlett's test of sphericity were significant (degrees of freedom = 66) and the KMO scores were above 0.7, indicating the data were suitable for a PCA (B cohort=0.75; K cohort=0.74).

The three extraction techniques (parallel analysis, scree plot, and Kaiser's criterion) revealed a three-component solution for each cohort individually, and both cohorts combined. Both orthogonal and oblique rotated solutions were assessed. The results were similar so the orthogonal rotation was reported. Only items with a component loading ≥0.3 were interpreted.

The results for three-component solutions explained 53.9% of the variance for the B Cohort, 55.0% of the variance for the K Cohort, and 53.9% of the variance for the cohorts combined. The results were similar for the B and K Cohorts individually and combined, and therefore results are presented for the combined sample. The Reactivity items loaded above the 0.3 threshold onto Component 1, the Approach items loaded onto Component 2, and the Persistence items loaded onto Component 3. The reduced set of SATI items used in LSIC showed good construct validity, with the three subscales emerging as clear distinct factors (Table 8.2).

Table 8.2. PCA for parent or carer reported SATI, B and K Cohorts combined aged 7.5-9 years (three-component solution)

	Component		
	1	**2**	**3**
1. Cries or complains loudly to a disappointment or failure (e.g. if <he/she> is not able to do something)	0.422		
2. Needs to be reminded to do homework *(Reverse scored)*	0.489		
3. Is shy with adults <he/she> doesn't know *(Reverse scored)*	0.509		
4. When angry, yells or snaps at others	0.453		
5. Goes back to the task at hand (chore, housework, etc.) after an interruption		0.489	
6. When meeting new children, acts shy or uncomfortable *(Reverse scored)*		0.542	
7. Sulks when <he/she>'s behaved badly (done the wrong thing)		0.478	
8. Does homework without anyone reminding <him/her>		0.462	
9. Seems uncomfortable when at someone's house for the first time *(Reverse scored)*			0.485
10. Gets very upset if anyone tells <him/her> off			0.477
11. Has difficulty completing assignments (homework, chores etc.) *(Reverse scored)*			0.563
12. Approaches children <his/her> age even when <he/she> doesn't know them			0.440

> **Key points**: The reduced set of SATI items used in LSIC showed good construct validity, with the three subscales emerging as clear distinct factors.

Internal consistency

The Approach and Reactivity subscales had acceptable internal consistency, for each cohort individually, and for the two cohorts combined (Table 8.3).

The Persistence subscale had an internal consistency score approaching acceptable, for each cohort individually, and for the two cohorts combined (Table 8.3). Three of the four Persistence items related to homework and had the greatest non-response rate, which may have compromised the consistency across the items.

Table 8.3. Internal consistency of SITQ subscales, by Cohort and overall

Cohort	α		
	Approach	Persistence	Reactivity
B Cohort	0.70 ✓	0.65 ~	0.72 ✓
K Cohort	0.71 ✓	0.67 ~	0.72 ✓
Both Cohorts	0.71 ✓	0.66 ~	0.71 ✓

X = unacceptable (alpha < 0.6). ~ = approaching acceptable (alpha 0.6 - < 0.7). ✓ = acceptable (alpha 0.7 - 0.9).

Key points: Internal consistency was acceptable for the Approach and Reactivity subscale, but only approaching acceptable for the Persistence subscale. This may be due to three of the four items relating to homework, which the child may not have received.

Recommendations

1. **Evidence of validity:** The SATI Approach and Reactivity subscales appear to be (psychometrically) valid and reliable measures of a child's temperament.
    1.1. However, we are unaware of any qualitative assessment of the meaning and relevance of the SATI items within the Aboriginal and Torres Strait Islander context. Further qualitative item validation research across contexts and language groups is required to accept the validity of the measure for use with Aboriginal and Torres Strait Islander children.[19,21]
    1.2. Accordingly, SATI data with this population must be interpreted with caution, and should be interpreted within context.
    1.3. Researchers using the SATI need to be explicit about the limitations of using this measure with Aboriginal and Torres Strait Islander peoples.
    1.4. It is not appropriate to compare the SATI between Aboriginal and Torres Strait Islander and non-Indigenous children; we cannot assume that the SATI is conceptually equivalent (measuring the same thing) in these populations.

2. **Exercise caution in using the Persistence subscale:** Although it appeared as a distinct construct, the internal consistency of the Persistence subscale was only approaching acceptable, and items from this subscale had the highest rate of non-response.

5. **Combining cohorts:** These results suggest that it is appropriate for data users to analyse data from the cohorts individually, or to combine the two cohorts to create a pooled sample of children of the same age range.

9. Brief Infant Toddler Social Emotional Assessment (BITSEA)

Background

Purpose of BITSEA

The Brief Infant Toddler Social Emotional Assessment (BITSEA) is a 42-item screening tool used to assess social and emotional development in children aged 12 to 36 months of age.[101] Parents or carers are asked about their perceptions of their child's social, emotional and behavioural problems and competencies. BITSEA is designed to assess if children are at risk of, or are experiencing, social-emotional behavioural problems, and if children are experiencing delays in their social-emotional competence.[101]

BITSEA can be self-administered or can be conducted as a part of an interview and takes between five and ten minutes to complete.[101,102] The BITSEA was developed from the broader Infant-Toddler Social and Emotional Assessment (ITSEA) in the United States. Items were selected for inclusion in the brief version (BITSEA) based on advice from mental health experts and the empirical evidence on items with high loading within the ITSEA.[101]

The refined BITSEA items covered seven main themes from the ITSEA (39 items total): Internalising Behaviour (8 items), Externalising Behaviour (6 items), Dysregulation, (8 items), Competence (7 items), Social Relatedness (3 items), Maladaptive Behaviour (3 items), and Atypical Behaviour (4 items). The remaining three BITSEA items (to form the total of 42) are from the ITSEA, but these items don't belong to a specific domain.[101 p. 146]

Parents or carers are asked questions about their child's behaviour in the past month, and asked to respond on a three-point scale, according to the options of: not true/rarely (0), somewhat true/sometime (1), and very true/often (2). Some items also include the response option of no opportunity (N). For some items, parents or carers are asked to provide an example or additional information about their child's behaviours.

Coding and scoring

The items included in the BITSEA cover two key domains, Problems (31 items) and Competency (11 items); accordingly, a Problems score and Competency score can be calculated. Scores can be compared to norms generated from a United States sample.[102] A high score on the Problem subscale (within the highest 25% of the normative sample) is intended to indicate that there is a possible problem with the child's behaviour.[102,103] A low score on the Competency subscale (lowest 15% of the normative sample) is intended to indicate that there is a possible deficit or delay in competency.[102,103] However, these scores are not indicative of a diagnosis, and participants are encouraged to seek follow-up testing.[102,103]

Total scores for the two subscales are not comparable or combinable, due to the different of items, and differences in interpretation (i.e. a high Problems score is linked with problem behaviour but a high Competency score is linked with higher competency).

Evidence on the validity of the measure for Aboriginal and/or Torres Strait Islander children

There is evidence of validity and reliability of the assessment within a United States population and with some other populations; however, to our knowledge there is no information on the validity specific to the Aboriginal and Torres Strait Islander population.

Briggs-Gowan et al. conducted a study on 1,237 toddlers aged 12-36 months to assess the reliability and validity of the BITSEA in the United Sates.[101] They demonstrated that BITSEA is a reliable and valid screening instrument for social-emotional and behavioural problems as well as delays in social-emotional competence within a United States population.[101] A later study conducted by Briggs-Gowan et al. concluded that the BITSEA is a valid screening instrument for identifying infants who display or

have emerging problem social-emotional behaviours.[104] This paper highlighted that the tool can be used by professionals in early detection and screening of young children.[104]

Items from BITSEA were used in early waves of LSAC. A shortened 34-item BITSEA scale was used in the P1 questionnaire in LSAC Wave 2 with the B Cohort (aged 2-3 years).[105,106] Parents or carers were asked the complete set of items from the Competence subscale (11 items), and they were asked a subset of items from the Problem subscales (23 out of 31 total items). Misson et al. conducted a factor analysis of the 34 items used in LSAC. They found that while the items were intended to map onto two components (problem subscale and competence subscale), the items mapped onto three components, which were labelled as Externalising Behaviours, Internalising Behaviours, and Social Competence.[107] Three remaining items from the BITSEA Problems subscales that were included in LSAC were classified as 'other types of problems'.[105,106] The components identified by the factor analysis are aligned with the key areas which BITSEA aims to measure.

Marmor et al. examined BITSEA in relation to SEWB in LSIC; she observed a moderate correlation between the BITSEA Problem score and Total Difficulties Score.[57] No interpretation was provided for this result.

Results: assessment of BITSEA in LSIC

Acceptability

Overall, non-response was very low, with an average of 10.0 item non-responses per 100 surveys. The item apa521 had the highest extent of item non-response rate, 2.0 per 100 surveys. Although respondents had the option to refuse an item, no respondent refused to answer a BITSEA question. One respondent did 'not answer' any BITSEA question.

> **Key points**: Overall, the extent of non-response to BITSEA items was low for parents or carers in this sample, indicating likely acceptability of the items.

Construct validity

Overall, inter-item correlations were low; of the possible 561 inter-item correlations, only 7 were above the 0.3 threshold. The results from the Bartlett's test of sphericity were significant (degrees of freedom = 561) and the KMO score of 0.80 indicated that the data were suitable for a PCA; as such, all items were included in the analysis.

The three extraction techniques used to identify the number of components revealed a six-component solution (parallel analysis and scree plot) and an 11-component solution (Kaiser's criterion). Only the six-component solution was investigated, as 11 components was far beyond the hypothesised two-subscale structure. The orthogonal rotated solution was assessed and reported. Only items with a component loading ≥ 0.3 were interpreted.

The six-component solution explained 36.6% of the variance (Table 9.1). Components 1, 2, and 4 included items from the Problems subscale; Component 3 included two items from the Competency subscale; and Components 5 and 6 included a combination of the Problems and Competency items.

The components identified in LSIC had some similarity with the three factors identified in analysis of data from LSAC, of 'Internalising' behaviours, 'Externalising' behaviours, and 'Social Competence'.[107] In LSIC, findings were consistent with Component 1 measuring 'Internalising' behaviour, Component 2 'Externalising behaviour' and Component 3 'Social Competence'. However, there were slightly fewer LSIC items loading above the 0.3 threshold on each component, compared to the LSAC item loadings.

Items that loaded above the 0.3 threshold onto Component 4 were related to sleep, but as there were only two items, these were not assessed for internal consistency. The authors did not identify a clear theme underlying the items that loaded above 0.3 onto Components 5 or 6. Therefore, these items were excluded from the internal consistency analyses below.

> **Key points**: Results were not consistent with the hypothesised two-factor BITSEA structure. The results were somewhat consistent with the a three-component solution identified in LSAC, measuring Internalising Behaviours, Externalising Behaviours, and Social Competence; however, the fit was not neat in LSIC.

Table 9.1. PCA for the parent or carer reported BITSEA, B Cohort Wave 2 (six-component solution)

	Component					
	1	2	3	4	5	6
apa52	0.324					
apa53	0.384					
apa58	0.326					
apa59	0.308					
apa510	0.319					
apa511	0.385					
apa512	0.308					
apa527		0.393				
apa528		0.363				
apa529		0.518				
apa530		0.448				
apa51			0.304			
apa54			0.415			
apa55			0.467			
apa56			0.353			
apa57			0.392			
apa56				0.601		
apa57				0.590		
apa55					0.452	
apa56					0.389	
apa57					0.430	
apa59					0.325	
apa510						0.401
apa511						0.344
apa513						0.546
apa514						-0.317

Internal consistency

Internal consistency for Component 1 (Internalising), Component 2 (Externalising) and Component 3 (Social Competence) were tested (Table 9.2). Despite the items loading onto distinct components, the internal consistency was below acceptable for each measure. The low internal consistency might be partly explained by the young age of study children at the time of measurement.

Table 9.2. Internal consistency of the three possible BITSEA subscales

Component	α
Internalising	0.63 ~
Externalising	0.67 ~
Competence	0.58 X

X = unacceptable (alpha < 0.6). ~ = approaching acceptable (alpha 0.6 - < 0.7). ✔ = acceptable (alpha 0.7 - 0.9).

> ***Key points***: Internal consistency was not acceptable for any of the components.

Recommendations

1. **Lack of evidence of validity:** The BITSEA, as administered in LSIC, did not perform well as a measure of social and emotional delays in Aboriginal and Torres Strait Islander children. The PCA revealed three components that were similar to those observed in the analysis of BITSEA in LSAC; however, internal consistency of the components was not considered acceptable in LSIC. It is recommended that data users identify items relevant to their research question and conduct independent validity and reliability checks.

   1.1. It is not appropriate to compare the BITSEA between Aboriginal and Torres Strait Islander and non-Indigenous children; we cannot assume that the BITSEA is conceptually equivalent (measuring the same thing) in these populations.

11. References

1.   Thurber KA, Banks E, Banwell C. Cohort Profile: Footprints in Time, the Australian Longitudinal Study of Indigenous Children. *International Journal of Epidemiology* 2014; **44**(3): 789-800.

2.   Thurber KA, Olsen A, Guthrie J, et al. 'Telling our story... Creating our own history': caregivers' reasons for participating in an Australian longitudinal study of Indigenous children. *International Journal for Equity in Health* 2018; **17**(1): 143.

3.   World Health Organisation. Mental Health: a state of well-being. 2014. https://www.who.int/features/factfiles/mental_health/en/ (accessed 4 May 2019).

4.   National Aboriginal Health Strategy Working Party. A National Aboriginal Health Strategy. Canberra: Commonwealth of Australia; 1989.

5.   Williamson A, Redman S, Dadds M, et al. Acceptability of an emotional and behavioural screening tool for children in Aboriginal Community Controlled Health Services in urban NSW. *Australian and New Zealand Journal of Psychiatry* 2010; **44**(10): 894-900.

6.   Department of Health and Ageing. Development of a Renewed Aboriginal and Torres Strait Islander Social and Emotional Framework- Discussion Paper. In: Department of Health and Ageing, editor. Canberra: Commonwealth of Australia; 2013.

7.   Jorm AF, Bourchier SJ, Cvetkovski S, Stewart G. Mental health of Indigenous Australians: a review of findings from community surveys. *The Medical Journal of Australia* 2012; **196**(2): 118-21.

8.   Priest NC, Paradies YC, Gunthorpe W, Cairney SJ, Sayers SM. Racism as a determinant of social and emotional wellbeing for Aboriginal Australian youth. *Medical Journal of Australia* 2011; **194**(10): 546-50.

9.   Ypinazar VA, Margolis SA, Haswell-Elkins M, Tsey K. Indigenous Australians' understandings regarding mental health and disorders. *Australian and New Zealand Journal of Psychiatry* 2007; **41**(6): 467-78.

10.   Swan P, Raphael B. Ways Forward: National and Torres Strait Islander Mental Health Policy. National Consultancy Report. Canberra: Commonwealth of Australia, 1995.

11.   Dingwall KM, Cairney S. Psychological and cognitive assessment of Indigenous Australians. *Australian and New Zealand Journal of Psychiatry* 2010; **44**(1): 20-30.

12.   Williamson A, Andersen M, Redman S, et al. Measuring mental health in Indigenous young people: A review of the literature from 1998–2008. *Clinical Child Psychology and Psychiatry* 2013; **19**(2): 260-72.

13.   Newton D, Day A, Gillies C, Fernandez E. A review of Evidence-Based Evaluation of Measures for Assessing Social and Emotional Well-Being in Indigenous Australians. *Australian Psychologist* 2015; **50**(1): 40-50.

14.   Le Grande M, Ski CF, Thompson DR, et al. Social and emotional wellbeing assessment instruments for use with Indigenous Australians: A critical review. *Social Science and Medicine* 2017; **187**: 164-73.

15.   Dauphinais P, King J. Psychological assessment with American Indian children. *Applied and Preventive Psychology* 1992; **1**(2): 97-110.

16.   Kowal E, Gunthorpe W, Bailie RS. Measuring emotional and social wellbeing in Aboriginal and Torres Strait Islander populations: an analysis of a Negative Life Events Scale. *International journal for equity in health* 2007; **6**(1): 18.

17.   Prout S. Indigenous wellbeing frameworks in Australia and the quest for quantification. *Social Indicators Research* 2012; **109**(2): 317-36.

18.   Oesterheld JR, Haber J. Acceptability of the conners parent rating scale and child behavior checklist to Dakotan/Lakotan parents. *Journal of the American Academy of Child & Adolescent Psychiatry* 1997; **36**(1): 55-64.

19.   Demarchi C, Bohanna I, Baune BT, Clough AR. Detecting psychotic symptoms in Indigenous populations: a review of available assessment tools. *Schizophrenia Research* 2012; **139**(1-3): 136-43.

20.   Javo C, Rønning JA, Handegård BH, Rudmin FW. Cross-informant correlations on social competence and behavioral problems in Sami and Norwegian preadolescents. *European Child & Adolescent Psychiatry* 2009; **18**(3): 154-63.

21.   Pace TM, Robbins RR, Choney SK, Hill JS, Lacey K, Blair G. A cultural-contextual perspective on the validity of the MMPI-2 with American Indians. *Cultural Diversity and Ethnic Minority Psychology* 2006; **12**(2): 320-33.

22.   Kite E, Davy C. Using Indigenist and Indigenous methodologies to connect to deeper understandings of Aboriginal and Torres Strait Islander peoples' quality of life. *Health Promotion Journal of Australia* 2016; **26**(3): 191-4.

23.   Haswell-Elkins M, Sebasio T, Hunter E, Mar M. Challenges of measuring the mental health of Indigenous Australians: honouring ethical expectations and driving greater accuracy. *Australasian Psychiatry* 2007; **15**(Supplement): S29-S33.

24.   Department of Prime Minister and Cabinet. National Strategic Framework for Aboriginal and Torres Strait Islander Peoples' Mental Health and Social and Emotional Wellbeing. In: Department of Prime Minister and Cabinet, editor. Canberra: Commonwealth of Australia; 2017.

25.   Braun KL, Browne CV, Ka 'opua LS, Kim BJ, Mokuau N. Research on Indigenous Elders: From positivistic to decolonizing methodologies. *The Gerontologist* 2013; **54**(1): 117-26.

26.   Minkler M. Community-based research partnerships: challenges and opportunities. *Journal of Urban Health* 2005; **82**(2): ii3-ii12.

27.   Bartlett JG, Iwasaki Y, Gottlieb B, Hall D, Mannell R. Framework for Aboriginal-guided decolonizing research involving Metis and First Nations persons with diabetes. *Social Science & Medicine* 2007; **65**(11): 2371-82.

28.   Sherwood J. Colonisation–It's bad for your health: The context of Aboriginal health. *Contemporary Nurse* 2013; **46**(1): 28-40.

29.   Sherwood J. Do no harm: decolonising Aboriginal health research. Sydney: University of New South Wales; 2010.

30.   Grove N, Brough M, Canuto C, Dobson A. Aboriginal and Torres Strait Islander health research and the conduct of longitudinal studies: issues for debate. *Australian and New Zealand Journal of Public Health* 2003; **27**(6): 637-41.

31.   Smith LT. Decolonizing methodologies: Research and Indigenous peoples; 2005.

32.   Humphery K. Dirty questions: Indigenous health and 'Western research'. *Australian and New Zealand Journal of Public Health* 2001; **25**(3): 197-202.

33.   Field A. Discovering Statistics Using IBM SPSS Statistics- And Sex, Drugs and Rock 'N' Roll. 4th ed. Singapore: Sage 2013.

34.   Tabachnick BG, Fidell LS. Using Multivariate Statistics. 5th ed. Sydney: Pearson; 2007.

35.   Kline P. A Handbook of Psychological Testing. 2nd ed. London: Taylor & Frances/Routledge; 1999.

36.   Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment* 2003; **80**(1): 99-103.

37.   Tavakol M, Dennick R. Making sense of Cronbach's alpha. *International Journal of Medical Education* 2011; **2**: 53-5.

38.   McClelland DC. Motive dispositions: The merits of operant and respondent measures. *Review of Personality and Social Psychology* 1980; **1**: 10-41.

39.   Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. United States of America: Lawrence Erlbaum Associates; 1988.

40.   Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry* 2007; **52**(2): 121-8.

41.   Glascoe FP. Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews* 2005; **11**(3): 173-9.

42.   Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development* 2011; **82**(1): 405-32.

43.   Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry* 1997; **38**(5): 581-6.

44.   Goodman R. A modified version of the Rutter parent questionnaire including extra items on children's strengths: a research note. *Journal of Child Psychology and Psychiatry* 1994; **35**(8): 1483-94.

45.   Youth in Mind. Scoring the Strengths & Difficulties Questionnaire for age 4-17. . 2014. Available from: http://www.sdqinfo.com/py/sdqinfo/c0.py (accessed 28 Apr, 2019).

46.   Youth in Mind. Scoring the Strengths & Difficulties Questionnaire for age 2-4. . 2014. Available from: http://www.sdqinfo.com/py/sdqinfo/c0.py (accessed 28 Apr, 2019).

47.   Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry* 1998; **7**(3): 125-30.

48.   Mellor D. Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist* 2005; **40**(3): 215-22.

49.   Goodman A, Lamping DL, Ploubidis GB. When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of Abnormal Child Psychology* 2010; **38**(8): 1179-91.

50.   Williamson A, McElduff P, Dadds M, et al. The Construct Validity of the Strengths and Difficulties Questionnaire for Aboriginal Children Living in Urban New South Wales, Australia. *Australian Psychologist* 2014; **49**(3): 163-70.

51.   Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 1999; **40**(5): 791-9.

52.   Keown LJ, Sanders MR, Franke N, Shepherd M. Te Whānau Pou Toru: A randomized controlled trial (RCT) of a culturally adapted low-intensity variant of the Triple P-Positive Parenting Program for Indigenous Māori families in New Zealand. *Prevention Science* 2018; **19**: 954-65.

53.   Bania EV, Eckhoff C, Kvernmo S. Not engaged in education, employment or training (NEET) in an Arctic sociocultural context: the NAAHS cohort study. *BMJ open* 2019; **9**(3): e023705.

54.   Young CRP. The Resilience of Urban Aboriginal Children and their Caregivers: University of Sydney 2019.

55.   Committee on Children with Disabilities. Developmental surveillance and screening of infants and young children. *Pediatrics* 2001; **108**(1): 192-5.

56.   Zubrick S, Lawrence D, De Maio J, Biddle N. Testing the reliability of a measure of Aboriginal children's mental health: an analysis based on the Western Australian Aboriginal Child Health Survey- Cat. No. 1351.0.55.011. In: Australian Bureau of Statistics, editor. Canberra: Commonwealth of Australia; 2006.

57.   Marmor A, Harley D. What promotes social and emotional wellbeing in Aboriginal and Torres Strait Islander children?: Lessons in measurement from the Longitudinal Study of Indigenous Children. *Family Matters* 2018; (100): 4-18.

58.   Williamson A, D'Este C, Clapham K, et al. What are the factors associated with good mental health among Aboriginal children in urban New South Wales, Australia? Phase I findings from the Study of Environment on Aboriginal Resilience and Child Health (SEARCH). *BMJ open* 2016; **6**(7): e011182.

59.   Williamson A, Skinner A, Falster K, Clapham K, Eades SJ, Banks E. Mental health-related emergency department presentations and hospital admissions in a cohort of urban Aboriginal children and adolescents in New South Wales, Australia: findings from SEARCH. *BMJ Open* 2018; **8**(11): e023544.

60.   Zubrick S, Lawrence D, Silburn S, et al. The Western Australian Aboriginal Child Health Survey (volume 1): the health of Aboriginal children and young people. Perth: Telethon Institute for Child Health Research. Perth: Telethoon Institute for Child Health 2004.

61.   Young C, Craig JC, Clapham K, Banks S, Williamson A. The prevalence and protective factors for resilience in adolescent Aboriginal Australians living in urban areas: a cross-sectional study. *Australian and New Zealand Journal of Public Health* 2018; **43**(1): 8-14.

62.   Lovett R. Indigenous Children's Resilience: The Role of Demographics, Relationships, Achievement and Culture. In: Walter M, Martin KL, Bodkin-Andrews G, eds. Indigenous Children Growing Up Strong. London: Palgrave Macmillan; 2017: 287-308.

63.   Williamson AB, Raphael B, Redman S, Daniels J, Eades SJ, Mayers N. Emerging themes in Aboriginal child and adolescent mental health: findings from a qualitative study in Sydney, New South Wales. *Medical Journal of Australia* 2010; **192**(10): 603-5.

64.   Bougie E, Arim RG, Kohen DE, Findlay LC. Validation of the 10-item Kessler psychological distress scale (K10) in the 2012 Aboriginal Peoples Survey. *Health reports* 2016; **27**(1): 3-10.

65.   Prince M. Measurement validity in cross-cultural comparative research. *Epidemiology and Psychiatric Sciences* 2008; **17**(3): 211-20.

66.   Sørlie T, Hansen KL, Friborg O. Do Norwegian Sami and non-indigenous individuals understand questions about mental health similarly? A SAMINOR 2 study. *International Journal of Circumpolar Health* 2018; **77**(1): 1481325.

67.   Stolk Y, Kaplan I, Szwarc J. Clinical use of the Kessler psychological distress scales with culturally diverse groups. *International Journal of Methods in Psychiatric Research* 2014; **23**(2): 161-83.

68.   Youth in Mind. Downloadable SDQs and related items. . 2014. Available from: http://www.sdqinfo.com/py/sdqinfo/b0.py (accessed 29 Apr, 2019).

69.   Northern California Training Academy. Mental Health Screening and Assessment Tools for Children- Literature Review. . Davis, California: Center for Human Services, University of California 2008.

70.   Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review. *Clinical Child and Family Psychology Review* 2010; **13**(3): 254-74.

71.   Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001; **40**(11): 1337-45.

72.   Dunstan DA, Todd AK. A method of assessing the resilience of whole communities of children: An example from rural Australia. *Child and Adolescent Psychiatry and Mental Health* 2012; **6**(1): 17.

73.   Thomas A, Cairney S, Gunthorpe W, Paradies Y, Sayers S. Strong Souls: development and validation of a culturally appropriate tool for assessment of social and emotional well-being in Indigenous youth. *Australian and New Zealand Journal of Psychiatry* 2010; **44**(1): 40-8.

74.   Bullen T. Strong Souls Review: MPH 5237 Clinical Measurement final assignment. 2017.

75.   Ungar M, Brown M, Liebenberg L, Othman R. Unique pathways to resilience across cultures. *Adolescence* 2007; **42**(166): 287.

76.   Westrupp EM, D'Esposito F, Freemantle J, Mensah FK, Nicholson JM. Health outcomes for Australian Aboriginal and Torres Strait Islander children born preterm, low birthweight or small for gestational age: A nationwide cohort study. *PloS one* 2019; **14**(2): e0212130.

77.   Thurber KA, Banwell C, Neeman T, et al. Understanding barriers to fruit and vegetable intake in the Australian Longitudinal Study of Indigenous Children: a mixed-methods approach. *Public Health Nutrition* 2017; **20**(5): 832-47.

78.   Dingwall KM, Cairney S. Detecting psychological symptoms related to substance use among Indigenous Australians. *Drug and Alcohol Review* 2011; **30**(1): 33-9.

79.   Jamieson LM, Paradies YC, Gunthorpe W, Cairney SJ, Sayers SM. Oral health and social and emotional well-being in a birth cohort of Aboriginal Australian young adults. *BMC Public Health* 2011; **11**(1): 656-67.

80.	Gresham FM, Elliott SN. SSIS Social–Emotional Learning Edition (SSIS-SEL) Manual. Bloomington, MN, USA: NCS Pearson; 2017.

81.	Greenberg MT, Weissberg RP, O'brien MU, et al. Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist* 2003; **58**(6-7): 466-74.

82.	Sherbow A, Kettler RJ, Elliott SN, Davies M, Dembitzer L. Using the SSIS assessments with Australian students: A comparative analysis of test psychometrics to the US normative sample. *School Psychology International* 2015; **36**(3): 313-21.

83.	Community-University Partnership for the Study of Children, Youth, and Families. Review of the Social Skills Rating System (SSRS). Edmonton, Alberta, Canada, 2011.

84.	Gresham FM, Elliott SN, Vance MJ, Cook CR. Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly* 2011; **26**(1): 27.

85.	McLean DL. Book Review: Social Skills Rating System (SSIS). *Journal of Psychoeducational Assessment* 1992; **10**(2): 196–205.

86.	James W. Crosby, Houston S. Test Review- Social Skills Improvement System (SSIS). *Journal of Psychoeducational Assessment* 2011; **29**(3): 292– 6.

87.	Gresham FM, Elliott SN. Social skills improvement system: Rating Scales Manual. Bloomington, MN, United States: PsychCorp Pearson; 2008.

88.	Collabpprative for Academic SaEL. CASEL Guide: Effective Social and Emotional Learning Programs. Chicago, IL, United States Collaborative for Academic, Social, and Emotional Learning, 2015.

89.	Davies M, Cooper G, Kettler RJ, Elliott SN. Developing social skills of students with additional needs within the context of the Australian Curriculum. *Australian Journal of Special Education* 2015; **39**(1): 37-55.

90.	Kettler RJ, Elliott SN, Davies M, Griffin P. Testing a multi-stage screening system: Predicting performance on Australia's national achievement test using teachers' ratings of academic and social behaviors. *School Psychology International* 2012; **33**(1): 93-111.

91.	Kaplun C, Knight J, Grace R, et al. Gudaga goes to school study: methods used in understanding school transitions and early education experiences of an urban Aboriginal cohort. *Educational Studies* 2016; **42**(1): 54-71.

92.	Rothbart MK, Bates JE. Temperament. In: Damon W, ed. Handbook of Child Psychology: Social Emotional and Personality Development. 5 ed. New York: Wiley; 1998.

93.	Little K, Sanson A, Zubrick SR. Do individual differences in temperament matter for Indigenous children?: The structure and function of temperament in Footprints in Time. *Family Matters* 2012; (91): 92-105.

94.	Sanson A, Hemphill SA, Smart D. Connections between temperament and social development: A review. *Social Development* 2004; **13**(1): 142-70.

95.	Kinchin I, Doran CM, McCalman J, et al. Delivering an empowerment intervention to a remote Indigenous child safety workforce: Its economic cost from an agency perspective. *Evaluation and Program Planning* 2017; **64**: 85-9.

96.	Kefalianos E, Onslow M, Ukoumunne OC, Block S, Reilly S. Temperament and early stuttering development: Cross-sectional findings from a community cohort. *Journal of Speech, Language, and Hearing Research* 2017; **60**(4): 772-84.

97.	Department of Social Services. Data User Guide- Footprints in Time: The Longitudinal Study of Indigenous Children, Release 9 (Waves 1-9). In: Department of Social Services, editor. Canberra: Commonwealth of Australia; 2018.

98.	Prior M, Sanson A, Smart D, Oberklaid F. Pathways from infancy to adolescence: Australian Temperament Project 1983–2000. In: Australian Institute of Family Studies, editor. Melbourne, Australia: Commonwealth of Australia; 2000.

99.	McClowry SG, Halverson CF, Sanson A. A re-examination of the validity and reliability of the School-Age Temperament Inventory. *Nursing Research* 2003; **52**(3): 176-82.

100.   McClowry SG. The development of the school-age temperament inventory. *Merrill-Palmer Quarterly* 1995; **41**(3): 271-85.

101.   Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV. The Brief Infant-Toddler Social and Emotional Assessment: screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology* 2004; **29**(2): 143-55.

102.   Community-University Partnership for the Study of Children, Youth, and Families, . Review of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). Edmonton, Alberta, Canada: Community-University Partnership, 2011.

103.   Smart D. Wellbeing of children and young people. Sydney: N.S.W.: Department of Family and Community Services, 2015.

104.   Briggs-Gowan MJ, Carter AS, McCarthy K, Augustyn M, Caronna E, Clark R. Clinical validity of a brief measure of early childhood social–emotional/behavioral problems. *Journal of Pediatric Psychology* 2013; **38**(5): 577-87.

105.   Smart D. How young children are faring: Behaviour problems and competencies. Melbourne: The Australian Institute of Family Studies, 2011.

106.   Baxter J. The family circumstances and wellbeing of Indigenous and non-Indigenous children. Melbourne: Australian Institute of Family Studies, 2013.

107.   Misson S, Sanson A, Berthelsen D, et al. Tracking children's development over time: The longitudinal study of Australian children outcome indices, waves 2 and 3. Melbourne: Australian Institute of Family Studies, 2011.