



Australian
National
University

TECHNICAL REPORT: MEASURING ADULT MENTAL HEALTH, PSYCHOLOGICAL DISTRESS, AND SOCIAL AND EMOTIONAL WELLBEING IN THE LONGITUDINAL STUDY OF INDIGENOUS CHILDREN

Due date: 28 May 2019

Katherine Thurber, Jennie Walker, Maggie Walter, Terry Dunbar, Jill Guthrie, Phil Batterham, Alison Callear, Alice Richardson, Lyndall Strazdins, Kate Doery, Ben Edwards, Raymond Lovett

Acknowledgements

This Report uses unit record data from *Footprints in Time*, the Longitudinal Study of Indigenous Children (LSIC), conducted by the Australian Government Department of Social Services (DSS). The findings and views reported in this paper, however, are those of the authors and should not be attributed to the Australian Government, DSS, or any of DSS' contractors or partners.

We acknowledge and thank the participants, and the leaders and Elders of their communities, who make the Study possible. We also acknowledge and thank the LSIC Research Administration Officers (RAOs), for their work in conducting interviews, and for sharing their insight with us in this project.

The authors acknowledge the contributions of: the LSIC Steering Committee; Fiona Skelton, Tracey Bullen, and Helene Shin from DSS; and Nicholas Biddle and Professor Matthew Gray from the ANU Centre for Social Research Methods.

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	1
The Longitudinal Study of Indigenous Children (LSIC)	1
Mental health and social and emotional wellbeing among Aboriginal and Torres Strait Islander peoples.....	1
The purpose of this report	3
The structure of this report	3
2. Overview of adult mental health-related measures used in LSIC.....	4
3. Methods.....	5
Research methodology	5
General analytical approach	5
Ethics.....	8
4. Kessler Psychological Distress Scale.....	9
Background	9
Purpose of the Kessler Psychological Distress Scale	9
Evidence on the validity of the measure for Aboriginal and/or Torres Strait Islander adults.....	10
Modifications between original scale and scale used in LSIC	12
Results: assessment of K10 and K5 in LSIC	12
Acceptability.....	12
Construct validity	13
Internal consistency	14
Agreement between the K10 and the K5 in predicting level of psychological distress.....	16
Convergent validity	16
Criterion validity.....	17
Recommendations	21
Recommendations for data users.....	21
Recommendations for future data collections	21
5. The Strong Souls Index.....	23
Background	23
Purpose of the Strong Souls Index	23
Modifications between original scale and scale used in LSIC	23
Results: assessment of Strong Souls Distress items in LSIC	28
Acceptability.....	28
Construct validity	28
Internal consistency	30
Convergent validity	30

Correlation between Strong Souls Distress and Kessler scores.....	32
Results: assessment of Strong Souls Strengths items in LSIC	32
Acceptability.....	32
Construct validity	33
Internal consistency	35
Convergent validity	35
Discriminant validity	36
Recommendations	36
Recommendations for data users	36
6. References	37
Appendices.....	41
Appendix A. K10 and K5 Confirmatory Factor Analysis	41
K10	41
K5	41
Appendix B. Strong Souls Distress at Wave 8	42
Construct validity	42
Internal consistency	42

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) and a parent or carer (P1, often the child's mother). Surveys are additionally conducted with a secondary parent or carer (P2, often the child's father) and with the child's teacher, at some waves. 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 peoples

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 Australians, it cannot be assumed that these measures are valid for use with Aboriginal and Torres Strait Islander peoples, 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}

While LSIC is primarily designed as a cohort of Aboriginal and Torres Strait Islander children, by design, it includes a cohort of at least an equal number of parents or carers of Aboriginal and Torres Strait Islander children: a parent or carer interviewed about the study child (Parent 1 or P1) and a second parent or carer (Parent 2 or P2). Parent 1 is the parent or carer interviewed about the study child. The P1 may be the study child's birth mother, a non-birth parent, another family member, or a foster carer. The study aimed to continue interviewing the same parent or carer but over time P1s may have changed if the original P1 was unavailable. Across waves, 95-98% of the P1s are female, and 81-85% identify as Aboriginal and/or Torres Strait Islander. In this Report, we focus on the P1 only, and do not analyse data from the P2. The results presented in this Report include all P1s, regardless of gender or Indigenous status, in order to reflect the whole cohort rather than a subset. Therefore, the findings presented in this report are representative of all parent or carer caregivers in LSIC. Further research could examine differences between Aboriginal and Torres Strait Islander P1s and non-Indigenous P1s, and differences by age group or gender, including data from P2s (where available).

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 P1s 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 adults. This Report assesses the appropriateness of measures developed for other populations for use with Aboriginal and Torres Strait Islander adults, and also assesses the appropriateness of a measure developed with and for Aboriginal and Torres Strait Islander youth (Strong Souls Index).

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 the measures in Chapters 4-5, along with recommendations for data users.

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

Table 2.1 provides an overview of the measures relevant to adult mental health in LSIC. This includes the Kessler and Strong Souls scales. These are intended to measure psychological distress and resilience. LSIC includes the entire set of Kessler items, and a selection of items from Strong Souls. The measures are designed for use with adolescents and/or adults; to our knowledge, an exact age range for use is not specified.

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

Measure	Intended construct	Full or partial measure
Kessler-10	Psychological distress	Full
Kessler-5	Psychological distress	Full
Strong Souls: Distress	Psychological distress	Partial
Strong Souls: Strengths	Resilience	Partial

Table 2.2 provides details on the adult mental health-related measures used in LSIC, according to number of items, variable names in the dataset, and the waves and cohorts at which the measures were administered. The Strong Souls measures are administered most frequently; the Kessler items are asked for the first time in Wave 10, and are scheduled for inclusion biennially (Wave 12, Wave 14). All measures are self-reported by the adult.

Table 2.2. Adult mental health-related measures in LSIC

Measure	# items	Variable names	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
			% female	98	98	98	98	98	97	97	96	95
		% Indigenous	85	85	84	83	83	82	82	82	82	81
		Median age (years)	28	30	31	32	33	34	35	37	38	38
Kessler-10	10	asw18_1 to asw18_10										✓
Kessler-5	5	asw18_2, asw18_4, asw18_5, asw18_8, asw18_9										✓
Strong Souls: Strengths (resilience)	12-13	ass1_a to ass1_m	✓	(N)	(N)	✓	(N)	(N)	(N)	✓	(N)	
Strong Souls: Distress (anxiety, depression)	7	asw2 to asw8	✓	✓	✓	✓		✓		✓		✓

Questions are asked of all adults unless indicated as new P1 only (N). Note: this document focuses on mental health measures of the P1; some mental health-related data are also collected from the P2. The approximate percentage of P1s that are female and Indigenous, and the median age (in years) of P1s is presented for each wave.

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.

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.

This project is led and governed by Aboriginal 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 non-Aboriginal 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.

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 adults, 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. Non-response across items and measures was summed and presented as a rate per 100 surveys. A greater extent of non-response was interpreted as evidence of potential item unsuitability.

Non-response was examined as a proxy for acceptability. We note that non-response may have been affected by survey mode; that is, the completion of the items face-to-face compared to online might have restricted participants’ ability to skip items (non-response). This was evidenced when comparing

non-response on the Strengths and Difficulties Questionnaire (SDQ) between P1s and teachers (for details, see *Technical report: Measuring child mental health, psychological distress, and social and emotional wellbeing in the Longitudinal Study of Indigenous Children*). There was a higher extent of non-response in teacher-reported measures compared to P1-reported measures. This may reflect that the majority of teachers self-completed the SDQ on paper which meant they were freely able to skip any items. Acceptability of items was used with more direct measures to inform and provide strength to recommendations.

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. 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 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 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 the outcome measure for use in assessing the convergent/divergent validity of adult mental health-related measures. Respondents were asked to indicate if they have had clinical depression and/or anxiety in the previous 12 months. Where relevant, correlations between different measures of adult 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 below outlines potential outcome measures for use in assessing the convergent validity of adult mental health scales.

Table 3.1. Potential outcome measures for validity assessment

Outcome measure	# items	Variable names	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11
		% female	98	98	98	98	98	97	97	96	95	94	
		% Indigenous	85	85	84	83	83	82	82	82	82	81	
		Median age (years)	28	30	31	32	33	34	35	37	38	38	
Primary outcome													
Clinical depression and/or anxiety	1	aoc2_18	✓	✓	✓	✓	✓	(B)	✓		✓		✓

Questions are asked of all P1s unless indicated as B cohort only (B), K cohort only (K), or new P1s only (N). The prefix corresponding to study wave is removed from variable names. The approximate percentage of P1s that are female and age range (in years) of P1s is presented for each wave. * Wave 11 beta data on mental health-related diagnoses are provided for the purpose of this report.

Criterion validity: The performance of the Kessler versions was assessed against a self-reported measure of clinical depression and/or anxiety 12 months later (item included Table 3.1). Ideally the performance of the Kessler versions would be measured against a ‘gold standard’, but in the absence of a ‘gold standard’ the self-reported measure was used as it is the best available measure. Criterion validity was measured by sensitivity, specificity and the positive predictive value (PPV).

The *sensitivity* of the Kessler versions to identify respondents who reported clinical depression and/or anxiety 12 months later was tested. A desired range for the sensitivity is 70-80%.⁴⁰ In the current report this means 70-80% of respondents who self-reported clinical depression and/or anxiety at Wave 11 reported a high level psychological distress at Wave 10. The *specificity* of the Kessler versions to correctly identify respondents who reported no clinical depression and/or anxiety 12 months later was tested. A desired range for the specificity is 80%. In the current report this means 80% of respondents who reported no clinical depression and/or anxiety at Wave 11 had a low level of psychological distress at Wave 10. The PPV was tested to assess how well the Kessler versions predicted self-reporting of clinical depression and/or anxiety 12 months later. There is no desired range of the PPV as it depends on the number of respondents self-reporting clinical depression and/or anxiety at Wave 11.

We calculated Receiver Operating Characteristics (ROCs) for the Kessler versions. 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).⁴¹ The aim of this analysis was to identify cut-off points for the Kessler versions 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.

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. Kessler Psychological Distress Scale

Background

Purpose of the Kessler Psychological Distress Scale

The Kessler Psychological Distress Scale is designed as a global screening instrument to identify generalised psychological distress, based on items about anxiety and depressive symptoms.⁴³⁻⁴⁷ The Kessler Psychological Distress Scale was not designed to be a tool for identifying risk factors for specific mental illnesses or as a diagnostic instrument.

The Kessler Psychological Distress Scale was developed by Professor Kessler and colleagues in the 1990s to be used in the United States National Health Interview Survey. The pilot version consisted of 45 items from 16 domains including depression, anxiety, worry, arousal, fatigue, and thoughts of death.^{43,45,48} Following the initial mail out pilot, the measure was reduced to 32 items, which were then tested through a telephone survey.^{43,45} Results from the telephone survey resulted in a 10-item measure, the Kessler-10 (K10), and a shorter 6-item measure the Kessler-6 (K6).^{43,45}

The K10 asks a series of 10 items about the respondent's experience of anxiety or depressive symptoms in the previous 4 weeks or 30 days: *'In the last 30 days, about how often did you feel...'*. The Kessler questionnaire can be self-completed by the participant themselves or can be interviewer-administered; it is designed to be completed in between 2 and 3 minutes. A refined 5-item measure, the Kessler 5 (K5), was developed for use in the National Aboriginal and Torres Strait Islander Health Survey (NATSIHS) in 2004. It was developed in collaboration with Kessler, state and territory health authorities, and Aboriginal and Torres Strait Islander people, to respond to concerns about the appropriateness of the K10. The K5 includes five of the six items from the K6; it does not include the last item about being *'worthless'* as this was considered to be inappropriate to ask Aboriginal and Torres Strait Islander peoples.^{13,46-48} The K5 also includes several minor wording changes from the K6: the prompt *'in the past 4 weeks'* was changed to *'in the last 4 weeks'*; *'hopeless'* was changed to *'without hope'*; and, *'restless or fidgety'* was changed to *'restless or jumpy'*. A comparison of the K10, K6, and K5 items is provided in Table 4.1.

Table 4.1. Prompt and items in the K10, K6, and K5

K10 item	K6	K5	K5/LSIC adaptations
During the past 30 days, about how often did you feel...?		In the last 4 weeks, about how often did you feel...?	
1. Tired for no reason			
2. Nervous	✓	✓	
3. So nervous that nothing could calm you down			
4. Hopeless	✓	✓	'without hope'
5. Restless or fidgety	✓	✓	'restless or jumpy'
6. So restless you could not sit still			
7. Depressed			
8. So sad nothing could cheer you up	✓	✓	#8/9 swapped in LSIC
9. That everything was an effort	✓	✓	#8/9 swapped in LSIC
10. Worthless	✓		

The response options for the Kessler items are: *'none of the time'* (1), *'a little of the time'* (2), *'some of the time'* (3), *'most of the time'* (4), and *'all of the time'* (5). The Kessler score can only be calculated when there are responses to all of the items in the scale.⁴⁶ For respondents with complete data on the items, responses to all items are summed to generate a score ranging from 10-50 for the K10 and 5-25 for the K5. Higher scores are intended to reflect higher levels of psychological distress, and potential need for support.^{46,47}

There are different approaches to how K10 scores are grouped; the most common categorisation cut-offs used to categorise K10 and K5 scores in Australia are presented in Table 4.2.^{46,47} There is emerging research from other populations identifying clinically-validated Kessler cut-off scores.⁴⁹

Table 4.2. Standard categorisation of K10 and K5 in Australia

Level of psychological distress	K10	K5
Low	10-<16	5-<8
Moderate	16-<22	8-<12
High	22-<30	12-<15
Very high	30-50	15-<25

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

The K10 has extensively been used in surveys and in practice. The K10 was first used in Australia in the 1997 National Survey of Mental Health and Well Being. Since then, the K10 has been used in a range of predominantly non-Indigenous state and national level surveys, including surveys by the Australian Bureau of Statistics (ABS) and in the Household Income and Labour Dynamics in Australia (HILDA) survey.^{13,47,48} The K10 was also used in the 45 and Up Study in New South Wales, which includes a cohort of over 1,000 Aboriginal adults aged 45 years and over. The K5 has been used with Aboriginal and Torres Strait Islander peoples in several national surveys including the 2008 and 2014-15 National Aboriginal and Torres Strait Islander Social Survey (NATSISS; aged 15 years and over) and the 2004-05 and 2011-13 National Aboriginal and Torres Strait Islander Health Survey (NATSIHS; aged 18 years and over).

There is international evidence that the K10 has ‘adequate to good’ content validity, construct validity, concurrent validity, and test-retest reliability; it has been identified that further validation work is required to demonstrate its predictive validity and sensitivity to change. There is also evidence from the total Australian population that the K10 is a valid measure of psychological distress.⁴³

There is limited evidence on the cultural equivalence of the Kessler scales, including for Indigenous populations internationally.⁵⁰ One paper from the United States examined the validity of the K6 for use with 3,084 First Nations adults (aged 15 years and over) from two reservations. Before implementing the survey, the cultural appropriateness and relevance of measures were reviewed through focus groups with males, females, service providers, and Elders. No modifications to the K6 were suggested through this process. The authors found that the internal reliability was acceptable in the sample (0.83), and identified a one-factor solution with a satisfactory fit. They found that the K6 could function as a general indicator of the severity of an individual’s psychological distress.

Another paper evaluated the validity of the K10 in a survey of 17,089 First Nations peoples aged 15 years and over in Canada. The authors found that the K10 was ‘psychometrically sound’ as a measure of general psychological distress for First Nations people living off reserve, for Métis, and for Inuit, with a unidimensional factor structure, satisfactory internal reliability, and evidence of convergent validity. However, they note that, ‘Findings should not be interpreted to mean that the factorial structure and item parameters of the K10 are invariant across cultures or that the K10 items are interpreted the same way by First Nations people, Métis, Inuit, and the general Canadian population. Additional research, including qualitative studies, would contribute to understanding of ‘the most appropriate idioms of distress’ for Aboriginal peoples in Canada, and provide information on the cultural validity of the K10 ... Examination of the cultural validity of the K10 and of the construct validity of the scale using clinically meaningful mental health outcomes for Aboriginal peoples in Canada is

also warranted'. While the K5 was developed for Aboriginal and Torres Strait Islander peoples, there has been limited exploration of the validity of the measure for this population. An Australian Institute of Health and Welfare (AIHW) report evaluated the convergent validity of the K5 in the 2004-05 NATSIHS. They observed expected associations between K5 and measures including positive wellbeing, anger, number of life stressors, discrimination, removal from family, general health, mental and other health conditions, and financial stress. There identified associations between increasing K5 scores and increasing impact of distress on everyday life, including increasing number of days missed from work, decreasing ability to undertake normal activities, and increasing visits to health professionals.⁴⁷ Other psychometric properties of the K5 were not explored. A stakeholder workshop was held in 2006 to refine the measures of SEWB used in ABS surveys. The consensus was that the K5 was considered a good measure of psychological distress for Aboriginal and Torres Strait Islander peoples. They recommended considering use of the K10 in future surveys to enable comparability with other surveys, and/or with non-Indigenous Australians.⁴⁷

McNamara et al. compared the psychometric proprieties of the Kessler items (10- and 5-item versions) among 1,632 Aboriginal and 231,774 non-Indigenous participants of the 45 and Up Study.⁴⁶ All participants in this study were aged 45 years and over and based in New South Wales at the time of data collection. The 45 and Up Study included the original K10 version, and therefore the 5-item measure assessed in this study has a few small differences from the actual K5 (see Table 4.3). McNamara et al. found that the Kessler items were largely acceptable, according to item non-response. Despite concerns about the K10 item about 'worthlessness', the extent of non-response was not higher for this item compared to other items. Agreement between the K10 and K5 was high. Eighty-five percent of Aboriginal participants were classified in the same distress category according to the two measures; at most classifications were off by one category. The weighted Kappa statistic was 0.87, indicating a very high level of agreement. McNamara et al. also conducted a paired t-test of the K10 and the K5 amongst the Indigenous participants ($p=0.54$) which shows there was not a significant difference between scores for the two scales. When the authors conducted CFA, they found that for both the K10 and the K5, all items loaded onto a single factor of psychological distress. Internal consistency was acceptable for both measures, with a Cronbach's alpha of 0.93 for the K10 and 0.88 for the K5. They found that the K10 and K5 had nearly equivalent convergent validity, when measured in relation to the impact of emotional problems, recent treatment for anxiety or depression, and quality of life. The authors found similar psychometric properties for the K6.⁴⁶

The findings support use of both the K10 and the K5 for measuring psychological distress among Aboriginal adults aged 45 years and over.⁴⁶ They indicate that choice of the K10 versus the K5 would depend on the setting and the need for brevity and or comparability. They identify that further examine of the Kessler items is required to determine the validity for use with younger Aboriginal adults, for people speaking Aboriginal and/or Torres Strait Islander languages. Further, we currently lack evidence on aspects of validity including clinical predictive validity, test-retest reliability, and the ability to detect change over time.

The Kessler scale is generally used with adults, but there is some evidence from the total Australian population indicating its validity for use with children and adolescents (11-17 years). This study identified that briefer versions of the K10 (not necessarily the set of items in the K6 or the K5) might be equally useful in capturing general psychological distress and/or in predicting mental health conditions. While the psychometric properties have not been assessed to our knowledge, the ABS surveys do use the K5 with Aboriginal and Torres Strait Islander adolescents aged 15 years and over and 18 years and over, in the NATSISS and NATSIHS, respectively.

Modifications between original scale and scale used in LSIC

An interviewer administered and/or self-completed K10 was completed by P1s at Wave 10. The K10 used in LSIC is a slightly modified version of the K10. LSIC included the K5 items, including the modifications from the original K10 items, and the remaining five K10 items (including the item about being 'worthless', which has previously been identified as inappropriate to ask). Further, K10 items 8 and 9 were provided in the reverse order within LSIC compared to the original K10 instrument. We recommend following the standard ordering if the full K10 is used in future waves of LSIC.

In some surveys where the K10 is interview-administered, 'skip rules' are put in place such that if respondents answer 'none of the time' to item 2, 5, or 7, the respondent is not asked the subsequent question, and the response is auto-filled to 'none of the time'. The skips are used as the subsequent question is considered redundant if participants have responded they do not feel the related emotion in the preceding questions. These skip rules have been employed in international studies, such as the US National Comorbidity Survey and the 2012 Aboriginal Peoples Survey, and in Australian surveys with the total population including the 2013-14 National Health Survey and the Australian Child and Adolescent Survey of Mental Health and Wellbeing.

The skip system was not formally implemented in LSIC. However, in some cases this happened informally in interviewer-administered delivery of the K10. RAOs reported that when participants responded 'none of the time' to the relevant items, the RAOs sometimes felt uncomfortable asking the subsequent question and instead filled that response as 'none of the time'. While not the official protocol for LSIC, this is aligned with the skip protocol used in many other surveys, and could be implemented formally in LSIC for consistency.

Results: assessment of K10 and K5 in LSIC

A total K5 and K10 score was computed by summing responses to all five and all ten items, respectively; respondents missing any items had missing total Kessler scores. Total K10 scores ranged from 10 to 50 and K5 scores from 5 to 25, with greater scores indicating greater levels of psychological distress.

Acceptability

Kessler items were treated as non-response if the respondent answered don't know, or refused to answer the question. Responses of don't know were more common than refusals to answer, but, overall, the extent of non-response was fairly low. For the K10 items, there was a total of 182 item non-responses (69 items refused and 113 'don't know' responses), averaging 14.3 item non-responses per 100 surveys. For the K5 items, there was a total of 98 item non-responses, averaging 7.7 item non-responses per 100 surveys. The total number of respondents was 1,270; of these, 95.9% (n=1,218) had a K10 score (not missing any of the ten items) and 96.7% (n=1,228) had a K5 score (not missing any of the five items). The prevalence of complete K10 and K5 data was higher in this sample compared to previous findings,⁴⁶ where 82% had complete data for the K10 and 83.5% had complete data for the K5. The lesser extent of non-response to Kessler items in LSIC compared to the 45 and Up Study may reflect differences in the survey administration, given that most LSIC participants completed the Kessler items face-to-face, whereas all 45 and Up Study respondents completed the Kessler items on a paper form.

There were no items with a particularly high level of non-response. The most commonly missed items were: *without hope* (1.73 non-response per 100 surveys), *restless or jumpy* (1.73), *that everything was an effort* (1.65), and *worthless* (1.65) see Table 4.3. Given RAO concern and previous research identifying concern about asking people about feelings of worthlessness, we had hypothesised that

we would see a particularly high level of non-response to this item. However, consistent with previous research,⁴⁶ non-response was not materially higher for this item compared to other items.

Table 4.3. Total number of non-responses for the Kessler items by P1s (N=1270) at Wave 10

Kessler items	Item non-response			Non-response per 100 surveys
	Refused	Don't know	Total	
1. Tired out for no good reason	5	5	10	0.81
2. Nervous	7	13	20	1.63
3. So nervous that nothing could calm you down	9	11	20	1.63
4. Without hope	7	15	22	1.79
5. Restless or jumpy	8	14	22	1.79
6. So restless that you could not sit still	6	10	16	1.3
7. Depressed	6	7	13	1.06
8. That everything was an effort	7	14	21	1.71
9. So sad that nothing could cheer you up	6	11	17	1.38
10. Worthless	8	13	21	1.71
Any missing K10 items	69	113	182	14.82
Any missing K5 items	37	61	98	7.98

K5 items are highlighted in light blue.

Key points: The extent of non-response was fairly low for both the K10 and the K5, suggesting that the Kessler items are generally considered acceptable by participants. None of the Kessler items were considered particularly 'unacceptable', according to patterns of non-response. By nature of having fewer items, the extent of non-response was lesser for the K5 compared to the K10.

Construct validity

Principal components analysis

To date, one published study has evaluated the construct validity of the K10 and K5 with an Aboriginal and Torres Strait Islander sample. The authors tested the hypothesised one factor structure of each scale using CFA. To our knowledge, PCA has not previously been conducted to assess the K10 or K5 in an Aboriginal and Torres Strait Islander sample. We conducted an exploratory analysis using PCA to test whether the intended factor structure appeared naturally in the data. PCA was conducted for the K10 and K5 data for all P1s at Wave 10.

All Kessler items were correlated above the 0.3 threshold with majority correlating between $r = 0.4$ and $r = 0.7$. The results of the Bartlett test of sphericity and the KMO score indicated the data were suitable for a PCA (see Table 4.4).

Table 4.4. Results from the Bartlett test of sphericity and KMO Measure of Sampling Adequacy by Kessler version for P1s at Wave 10

	Bartlett test of sphericity χ^2 (p-value)	KMO
K10 <i>df</i> =45	7828.01 (<0.01)	0.93
K5 <i>df</i> =10	2953.65 (<0.01)	0.85

df = degrees of freedom.

The number of components for extraction was determined by the three-criterion approach (see Chapter 3 Methods). All three approaches recommended a one component solution be retained for the PCA for the K10 and K5. As only one component was retained, a rotated solution was not applicable. Only items with a component loading of >0.3 were interpreted and presented in the output. The PCA for the K10 explained 61.0% of the variance. All items loaded onto the component (loading >0.3) except for two items: *Tired out for no good reason* and *So restless that you could not sit*

still (see Table 4.5). The item *So restless that you could not sit still* had a loading very close to the 0.3 threshold at 0.299.

Table 4.5. Results from the PCA for the K10 for P1s at Wave 10

	Loading
	Component 1
1. Tired out for no good reason	
2. Nervous	0.309
3. So nervous that nothing could calm you down	0.315
4. Without hope	0.325
5. Restless or jumpy	0.315
6. So restless that you could not sit still	
7. Depressed	0.335
8. That everything was an effort	0.320
9. So sad that nothing could cheer you up	0.339
10. Worthless	0.325

Items with a loading <0.3 were excluded from the output.

The PCA for the K5 explained 66.5% of the variance and all items loaded onto one component above 0.4 (see Table 4.6). The K5 item loadings had a greater value compared to the K10, and hence a greater amount of variance was explained in the component.

Table 4.6. Results from the PCA for the K5 for P1s at Wave 10

	Loading
	Component 1
3. So nervous that nothing could calm you down	0.441
4. Without hope	0.459
5. Restless or jumpy	0.436
7. Depressed	0.460
8. That everything was an effort	0.440

Confirmatory factor analysis

We assessed the single factor structure of the K10 and K5 using a Confirmatory Factor Analysis. The results are presented in Appendix 1. In summary, the model was a poor fit of the data for the K10, but the model fit was improved when two pairs of error terms were correlated. However, the final K10 model still required improvement. The adjusted K5 model appeared to adequately fit the data, although one fit index was not considered acceptable.

Key points: The results showed that both the K10 and the K5 had a unidimensional factor structure, with all items loading on to a single construct of ‘psychological distress’. This is consistent with previous research in this population.⁴⁶ This provides support for the administration and use of the K10 or the K5 in LSIC. The K5 appears to be performing as well, if not better, than the K10, according to the amount of variance explained in each solution and the item loadings in the PCA, and an adequate-fitting adjusted model in the CFA.

Internal consistency

The internal consistency of K10 and K5 were assessed using Cronbach’s alpha (see Table 4.7). To test the robustness of findings to different measures of internal consistency, analysis was repeated using Raykov’s reliability. No difference in the results were observed using Cronbach’s alpha compared to Raykov’s reliability (data not shown); therefore, results from Cronbach’s alpha are presented in this Report. Both the K10 and the K5 had an internal consistency score above the ‘acceptable’ threshold.

The results were similar when the sample was restricted to P1s who identified as Aboriginal and/or Torres Strait Islander (almost 81.0% of respondents).

The internal consistency of both Kessler versions was evaluated across levels of remoteness, to test if the measure’s internal consistency was consistent across contexts. The alpha score appeared to be slightly lower in very remote regions compared to other areas, but remained above acceptable for participants living in all levels of remoteness, for both the K10 and the K5.

In all instances, Cronbach’s alpha for the K10 was ≥ 0.90 . This might reflect item redundancy within the K10, which would support removal of some items.

Table 4.7. Cronbach’s alpha for the K10 and K5 for P1s at Wave 10, overall and by level of remoteness

	α						
	Overall	Indigenous P1s only	Major city	Inner regional	Outer regional	Remote	Very remote
K10	0.92	0.93	0.93	0.93	0.93	0.92	0.90
	✓	✓	✓	✓	✓	✓	✓
K5	0.87	0.87	0.87	0.89	0.86	0.88	0.80
	✓	✓	✓	✓	✓	✓	✓

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

Key points: Both the K10 and the K5 have acceptable internal consistency for P1s in LSIC, consistent with previous research. Internal consistency remained acceptable even in very remote settings. There was some evidence of item redundancy within the K10.

On balance, we found that the K5 performed at least as well, if not better than the K10. This was evidenced by: the lesser extent of non-response in the K5 compared to K10; possible item redundancy within the K10, but not within the K5; the higher amount of variance explained in each PCA solution for the K5 compared to the K10; the higher item loadings for the K5 compared to the K10; the K5 had good fit for the adjusted model in the CFA; and, the K5 had acceptable internal consistency. Further, one K10 item (*tired out for no good reason*) did not load on to the component and one item had a component loading on the threshold (*so restless that you could not sit still*). The first item was identified by McNamara et al.⁴⁶ as having a low inter-item correlation. Further exploration of both of these items, particularly the former, may be warranted, including their appropriateness and whether a clarifying statement is required to facilitate consistent interpretation. In addition, RAOs raised concern about asking participants about feelings of *worthlessness*, consistent with feedback from previous consultations with Aboriginal and Torres Strait Islander stakeholders.^{13,46-48} While non-response was not materially higher for this item than for other items, and this item did load on to the component, it should not be assumed that this means it is appropriate to ask this question. RAOs voiced concern that asking this question might be harmful to participants, particularly those who were in a ‘fragile’ state. Further research is required to understand the appropriateness and acceptability of this item.

Even within the K5 the RAOs identified some items that sometimes required explanation or clarification, such as *without hope*, and *that everything was an effort*. Clarifying statements could be developed and piloted to support consistent explanation of the Kessler items across RAOs and across survey contexts. While this does not require any changes to the actual Kessler items, introduction of clarifying statements in subsequent waves of LSIC might compromise consistency across waves. That is, providing clarifying statements might change the way participants interpret and understand the Kessler items, compared to when the clarifying statements were not used. To explore the potential impact of implementing clarifying statements, responses with and without clarifying statements could

be compared within respondents, potentially in a separate pilot sample. An evaluation is recommended before implementing these statements.

Agreement between the K10 and the K5 in predicting level of psychological distress

Respondents were categorised as having low, moderate, high, or very high levels of psychological distress at Wave 10 according to K10 and K5 scores. Respondents with a K10 score of 10-15 were categorised as having a low level of psychological distress, 16-21 a moderate level, 22-29 a high level, and 30-50 very high level. Respondents with a K5 score of 5-7 were categorised as having a low level of psychological distress, 8-11 a moderate level, 12-14 a high level, and 15-25 very high level.

Overall agreement was calculated by summing the total number of respondents with category agreement (presented in bold) and dividing this by the total number of respondents ($N=1,218$). There was 88.2% agreement between the K10 and K5 categories (see Table 4.8); this is slightly higher than in previous research with this population (85.0% agreement overall). There was only one participant in LSIC (<1%) whose classification according to the K10 versus K5 varied by more than one category (K10: moderate; K5: very high).

Table 4.8. Number of P1s in each category of psychological distress, according to the K10 and the K5, at Wave 10

K10	K5				Total
	Low	Moderate	High	Very high	
Low	759	42	0	0	801
Moderate	40	163	3	0	206
High	0	46	79	8	133
Very high	0	1	4	73	78
Total	799	252	86	81	1,218

Key points: There was a high level of agreement in the categorisation of level of psychological distress using the K10 versus the K5. This means that the vast majority of participants are classified as having the same level of psychological distress whether the K10 or the K5 is used.

Convergent validity

This section included only respondents where the P1 was the same individual at Wave 10 and Wave 11, and where the respondent had a non-missing total Kessler score at Wave 10 and self-reported clinical depression and/or anxiety score at Wave 11. A total of 58 respondents who completed the clinical depression and/or anxiety measure at Wave 11 were different from the P1 who completed the Kessler measure at Wave 10. These 50 P1s were excluded from the analyses. We compared the percentage of P1s with clinical depression and/or anxiety across levels of psychological distress, according to the K10 and the K5. We also calculated odds ratios (ORs) and 95% Confidence Intervals (CIs) to test if the odds of self-reported clinical depression and/or anxiety were significantly associated with level of psychological distress.

We found that the percentage of P1s with self-reported clinical depression and/or anxiety increased with increasing level of psychological distress, as measured by the K10 and the K5. Of those with low psychological distress according to the K10 and the K5, 4.2% and 4.1% reported clinical depression and/or anxiety, respectively. This increased to 12.1% and 15.8% of those with moderate levels of psychological distress, to 27.8% and 24.7% of those with high levels of psychological distress, and to 32.6% and 36.5% of those with very high levels of psychological distress.

The odds of reporting clinical depression and/or anxiety were significantly higher among participants with higher (compared to low) levels of psychological distress, according to both the K10 and the K5.

Using the K10 categorisation, compared to those with low levels of psychological distress, those with moderate levels of psychological distress had >3 times the odds of having clinical depression and/or anxiety, those with high levels had >8 times the odds, and those with very high levels had >12 times the odds. Respective figures for the K5 were >4, >7, and >13.

Table 4.9. Number of respondents who reported clinical depression and/or anxiety at Wave 11 by psychological distress category and Kessler version

Level of psychological distress	No clinical depression or anxiety % (n)	Clinical depression and/or anxiety % (n)	Odds Ratio of having clinical depression and/or anxiety (95%CI)
K10			
Low (n=662)	95.8 (664)	4.2 (28)	1 (ref)
Moderate (n=175)	86.9 (152)	12.1 (23)	3.43 (1.92, 6.11)
High (n=108)	72.2 (78)	27.8 (30)	8.71 (4.94, 15.34)
Very high (n=59)	64.4 (38)	35.6 (21)	12.51 (6.51, 24.06)
K5			
Low (n=661)	95.9 (634)	4.1 (27)	1 (ref)
Moderate (n=216)	84.3 (182)	15.8 (34)	4.39 (2.58, 7.46)
High (n=73)	75.4 (55)	24.7 (18)	7.68 (3.98, 14.82)
Very high (n=63)	63.5 (40)	36.5 (23)	13.50 (7.11, 25.64)

Overall, our findings indicate good psychometric properties of both the K10 and the K5 in this sample of predominantly Aboriginal and Torres Strait Islander adults. For both the K10 and the K5: the extent of non-response was fairly low; there was support for the intended unidimensional factor structure; internal consistency was acceptable, even within very remote contexts; and, higher compared to lower levels of psychological distress were associated with significantly increased odds of mental health conditions. We observed a high level of agreement between the K10 and the K5, with 88.2% of participants being classified in the same level of psychological distress according to both measures. Convergent validity was equivalent for the two measures. Adding strength to our findings, the results from this sample are consistent with results from previous research with older Aboriginal adults in New South Wales.

Criterion validity

As shown in Table 4.10, when the Kessler cut-off score was at the lower bound of the moderate category, the sensitivity of the Kessler scales reached the desired range of 70-80%. This means that just over 70% of respondents who self-reported clinical depression and/or anxiety at Wave 11 were considered as having moderate to very high risk of psychological distress at Wave 10. However, the specificity fell below the desired 80% value, at 70%. This means 70% of adults who did not report clinical depression and/or anxiety at Wave 11 were classified as low risk for psychological distress. The PPV of the Kessler versions with score at the lower bound of the moderate category was 21%, indicating 21% of adults with a moderate to very high risk of psychological distress reported clinical depression and/or anxiety in the following 12 months. As such 79% of adults with a moderate to very high risk of psychological distress did not report clinical depression and/or anxiety in the following 12 months, suggesting the vast majority of adults who screen as moderate to high risk would not self-report clinical depression and/or anxiety 12 months later. The low PPV might partly be explained by the low percentage of respondents self-reporting clinical depression and/or anxiety.

As shown in Table 4.10, when the Kessler cut-off score was increased to the lower bound of the high category, the sensitivity of the Kessler scales was reduced well below the desired range of 70-80%. Half of adults who reported clinical depression and/or anxiety at Wave 11 had high or very high risk of psychological distress according to the K10, and approximately 40% had high or very high risk of psychological distress according to the K5. The specificity for both Kessler versions was close to 90%.

This means almost 90% of all adults who did not report clinical depression and/or anxiety at Wave 11 were identified as having low or moderate risk of psychological distress at Wave 10. The PPV of the Kessler versions with a score at the lower bound of the high category was 30%, indicating 30% of adults with a high or very high risk of psychological distress reported clinical depression and/or anxiety in the following 12 months. As such 70% of adults with a high or very high risk of psychological distress did not report clinical depression and/or anxiety in the following 12 months.

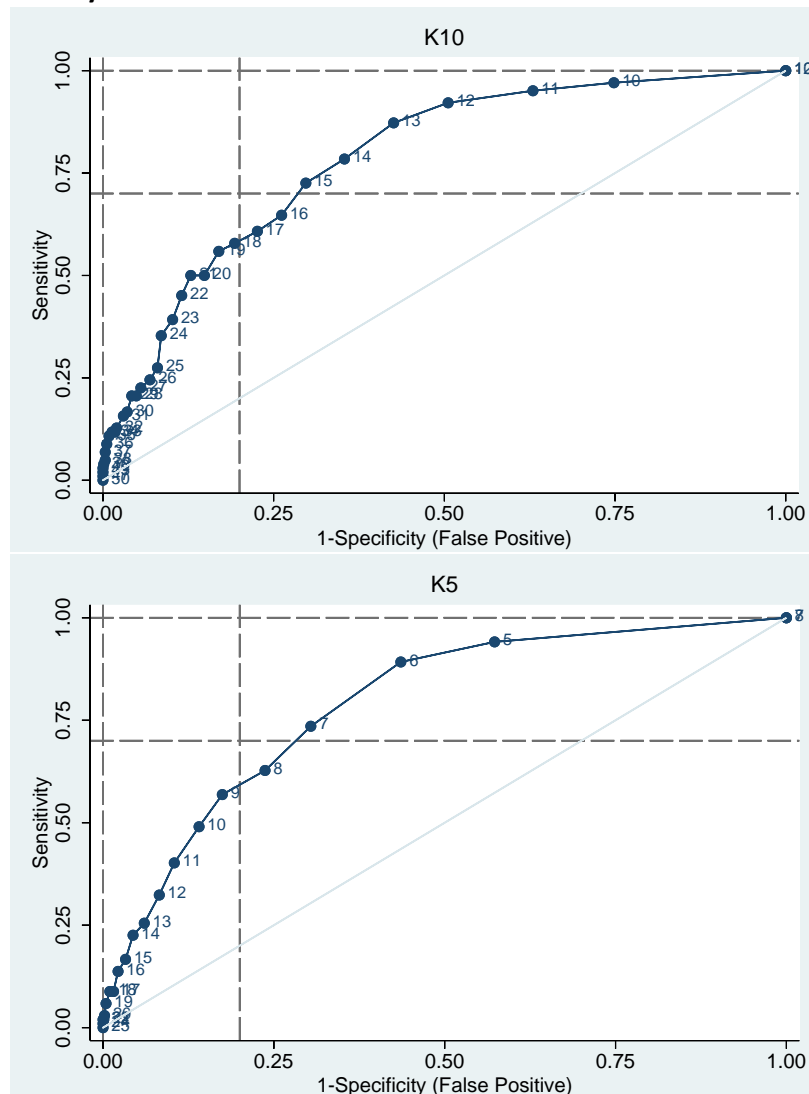
These findings indicate that while there was a strong association between the Kessler-10 and Kessler-5 at Wave 10 and self-reported clinical depression and/or anxiety at Wave 11, the Kessler was a poor predictor of clinical outcomes 12 months later.

Table 4.10. Sensitivity, specificity, and PPV of the K10 and K5 (dichotomised) in relation to self-reported clinical depression and/or anxiety

	Sensitivity (%)	Specificity (%)	PPV (%)
K10 (cut-off)			
'Low' vs. 'moderate to very high' (≥ 15)	72.6	70.3	21.6
'Low and moderate' vs 'high and very high' (≥ 22)	50.0	87.1	30.5
K5 (cut-off)			
'Low' vs. 'moderate to very high' (≥ 7)	73.5	69.6	21.3
'Low and moderate' vs 'high and very high' (≥ 12)	40.2	89.6	30.2

The ROC curves in Figure 1 demonstrate the trade-off between sensitivity and specificity. The results from the ROC curves show the optimal score to correctly identify at least 70% of respondents who reported clinical depression and or anxiety was at the lower bound of the moderate category. This cut-off resulted in a high false positive rate capturing just over 20.0% of respondents who reported clinical depression and/or anxiety in the following 12 months. Based on the current results it appears the Kessler versions administer in LSIC are a poor predictor of self-reported clinical depression and/or anxiety 12 months later.

Figure 1. ROC curve for Kessler score at Wave 10 in relation to self-reported clinical depression and anxiety at Wave 11



To the authors knowledge this was the first time a sensitivity and specificity analysis was conducted using data from an Aboriginal and Torres Strait Islander sample. While there was a strong association between the Kessler-10 and Kessler-5 at Wave 10 and self-reported clinical depression and/or anxiety at Wave 11, the Kessler scales were poor predictors of the outcome 12 months later. To correctly identify adults who self-reported clinical depression and/or anxiety at Wave 11, the Kessler cut-off was at the lower bound of the moderate category. This resulted in a high false positive rate. In contrast, to correctly identify adults who did not report clinical depression and/or anxiety at Wave 11, the Kessler cut-off was at the lower bound of the high category. This resulted in approximately half of respondents with clinical depression and/or anxiety at Wave 11 at high or very high risk of psychological distress, misclassifying the other half as low to moderate risk. The choice of cut-offs used would depend on the purpose of the research.⁵¹ It is worth noting that there is some evidence the Kessler versions do not capture distress related to less prevalent mental disorders as accurately as some other distress measures. While this does not mean it should not be used, distress related to disorders such as post-traumatic stress disorder, obsessive compulsive disorder, or social anxiety may not be accurately captured by the Kessler due to its focus on depression and generalised anxiety disorder symptoms.^{52,53}

It must also be noted that self-reported clinical depression and/or anxiety is likely to reflect considerable under-reporting of mental disorders, as rates of treatment seeking (and consequent

diagnosis) for mental health problems is very low particularly among Aboriginal and Torres Strait Islander people,⁵⁴ as is recognition of mental health problems.⁵⁵ Nonetheless, the findings suggest that the cut-offs on the Kessler scales are unlikely to be sufficiently accurate to merit use in a screening context.

Key points: We observed a strong association between level of psychological distress at Wave 10 and self-reported clinical depression and/or anxiety at Wave 11. The pattern of results was consistent when using the K10 and the K5 as the measure of psychological distress; this is consistent with previous research.⁴⁶ However, both Kessler versions had a low sensitivity and positive predictive value in relation to self-reported clinical depression and/or anxiety, according to the current high cut-off. Although there was a significant relationship between the Kessler versions and self-reported clinical depression and/or anxiety, the Kessler appears to be a poor predictor of respondents reporting clinical depression and/or anxiety 12 month later.

Based on our findings, we recommend consideration of use of the Kessler items with Study Children in LSIC as they approach adolescence. The K5 has been used with Aboriginal and Torres Strait Islander children as young as 15 years (in the NATSISS),^{56,57} and the K6 is currently being used in the Generation Next Study (in addition to the SDQ) for Aboriginal children aged 15 years and older. The Generation Next Study pilot tested the survey with a sample of Aboriginal adolescents to assess the ‘appropriateness and comprehension of the survey questions’ before finalising the survey. While detailed information is not available, inclusion of K6 in the final Generation Next Study survey may indicate that the K6 items were generally considered appropriate and understood by the pilot participants. In addition, LSAC, which includes a small cohort of Aboriginal and Torres Strait Islander youth, employed the K10 at Wave 8, when the participants were 18-19 year old. The K10 will continue to be used in LSAC biennially.

The LSIC RAOs commented that they generally thought it could be valuable to use the Kessler items with the study children, once they reached a suitable age. If included in future surveys, the RAOs suggested the Kessler items be self-completed by the study child, rather than interviewer-administered. Similarly, Aboriginal adolescents in the Generation Next Study will be self-completing the Kessler items on tablets.⁵⁸ A study with adolescents in the total Australian population also used self-completion rather than interviewer-administration of the Kessler items.⁵⁹

It has previously been argued the K10 should be considered for use over the K5, to enable comparability with other surveys. While comparability is beneficial, we do not consider that this benefit outweighs the benefits of using the K5 described above, or that it outweighs the concerns associated with using the full set of K10 items. First, use of the K5 does not preclude comparison with other studies that have employed the K10, given that the K5 is a subset of the K10. Where researchers or policymakers desire to compare Kessler scores in LSIC to scores in another sample that used the K10, the subset of K5 items from the relevant survey could be compared to the K5 items used in LSIC. Second, the primary purpose of LSIC is to provide evidence on the wellbeing of Aboriginal and Torres Strait Islander children and their families; the primary purpose of LSIC is not to provide a comparison point for other samples, including the non-Indigenous population. Accordingly, when selecting measures for inclusion in LSIC, the measure’s comparability should be a lesser priority compared to the measure’s appropriateness for use with the Aboriginal and Torres Strait Islander population. This approach has been maintained throughout the development and refinement of LSIC to date. Third, conceptual equivalence of the Kessler items cannot be assumed across populations. While the current research has identified sound psychometric properties of the K10 and K5 in this population, additional cross-cultural validation research would be required to understand if the constructs are culturally relevant and if the items are interpreted the same way in this population compared to in the non-Indigenous population (or other populations).^{50,60-62}

Recommendations

Recommendations for data users

1. The K10 and the K5 appear to be acceptable, internally reliable (across levels of remoteness), and valid for use as measures of psychological distress for adults in the Longitudinal Study of Indigenous children. On balance, our findings indicate that the K5 is performing better, or at least as well, as the K10.
2. The K10 and the K5 be used as a continuous scale in research models with lower scores indicating lower levels of psychological distress. The authors advise against using the Kessler versions as a screening tool to identify respondents with potential mental health issues.
3. It should not be assumed that Kessler scores are conceptually equivalent between different populations; therefore, it should not be assumed to be appropriate to compare Kessler scores between the Aboriginal and Torres Strait Islander and non-Indigenous population.
4. The Aboriginal and Torres Strait Islander sample in LSIC was selected using a non-probability sampling method. Selection was not random with families asked to participate from 11 sites across Australia. It is unknown how well the LSIC sample represents the total Aboriginal and Torres Strait Islander population; therefore, the results cannot be generalised to the entire population.

Further, it is not appropriate to compare Kessler scores between Aboriginal and Torres Strait Islander adults and non-Indigenous adults; we cannot assume that the Kessler versions are conceptually equivalent (measuring the same thing) in these populations.

5. When examining measures and/or outcomes at multiple time points, it is critical to ensure that the analysis is restricted to the caregivers who have remained constant across the time points of interest. That is, it is important to exclude participants where a different individual has responded as P1 at different time points. A variable is included at each Wave to indicate if the current P1 is the same as the previous Wave. However, at the time of writing this report, parents and carers (as well as teachers) did not have a unique identifier, making it difficult to match caregivers beyond one wave.

Recommendations for future data collections

1. These findings support the ongoing use of the Kessler items in LSIC. They provide support for use of the K5 rather than the K10 in future waves, given the K5's equivalent or even improved psychometric properties compared to the K10, with half the number of items. An additional benefit to using the K5 over the K10 is the reduction in respondent and interviewer burden. Given the multiple benefits, and lack of costs, to using the K5 rather than the K10, we recommend use of the K5 in future survey administrations.
2. If the full set of K10 items are retained in LSIC, we recommend that LSIC employs the same ordering as the original K10 items. We recommend LSIC considers employing a 'skip' protocol for K10 items 3, 6, and 8, as has been done in other face-to-face administrations of Kessler items.
3. We recommend the development of clarifying statements, in collaboration with RAOs and participants, to enable the consistent explanation of Kessler items across participants, contexts, and waves. Piloting and evaluation would be required before implementing these clarifying statements in future surveys.
4. We recommend consideration of use of the Kessler items with LSIC Study Children in subsequent waves. Before including these measures, however, we recommend piloting the measures with Aboriginal and Torres Strait Islander children in the relevant age range, and

their parent or carer, to ensure appropriateness. Consultation processes would determine the minimum age for administration of the Kessler items, the number of Kessler items included, and the mode of administration.

5. To support data users to analyse LSIC data longitudinally, we recommend that DSS provide a unique identifier to all P1s, P2s and teachers.

5. The Strong Souls Index

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. It was designed to include concepts relating to depression, anxiety, suicide risk, and resilience. A total pool of items was created and then refined through a review of the literature and ongoing consultative processes with Aboriginal community members and mental health experts. 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”.

The final pool of 34 Strong Souls items were tested in a pilot study with 361 Aboriginal participants (youth), alongside two existing measures, the K6 and the WASC-Y. 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. 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). It encompasses two subscales: SEWB and resilience.⁶³

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. 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. 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. An overall distress score can also be calculated by summing the scores for the anxiety, depression, and suicide risk subscales.

Modifications between original scale and scale used in LSIC

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. 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 final Strong Souls Index items and response options, and the Strong Souls items used in LSIC, are displayed in Table 5.1.

Of the 7 total Strong Souls: Distress items used in LSIC, 6 are from the final Strong Souls Index, 3 of which map on to the anxiety factor, and 1 of which maps on to the depression factor. The remaining 1 item in the LSIC Strong Souls: Distress measure is from the initial pool of Strong Souls items but is not included in the final Strong Souls Index.

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 (Wave 8) is the item about ‘*You know lots about white fella ways*’. This item was included in earlier waves of LSIC, and only asked of Aboriginal and/or Torres Strait Islander respondents. The item was removed

from the survey in response to RAO feedback: ‘Subsequent pilot testing of the resilience subscale items occurred in another longitudinal study and 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.’ Because this item was not included in all waves, we have not included this item in the table or in our analysis. 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. 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.

Items from the Strong Souls Index were administered to children in the K cohort at Wave 9, and children in the B and K cohort at Wave 12; Wave 12 collection was underway at the time of writing (for details, see *Technical report: Measuring child mental health, psychological distress, and social and emotional wellbeing in the Longitudinal Study of Indigenous Children*).



Table 5.1. Comparison of the Strong Souls Index final items against the Strong Souls Items used in LSIC

Strong Souls Index			LSIC Strong Souls items		
<p>Prompt items 1-16: These are some questions about things that sometimes happen to people. Please circle how often these things happened to you in the past few months. Prompt items 17-25: How much is this like you?</p>			<p>Strengths (Resilience) items Prompt: The next questions are about what helps you get through hard times. How much is this like you?</p>	<p>Distress items Prompt: The next questions are about how you feel about big worries, stress or sadness. Just think about the last few months. In the last 3 months (since... month) ...</p>	
Item	Response options	Factor			Response options
From final Strong Souls Index					
1. Have trouble sleeping?	Not much, sometimes, fair bit, lots of times	Depression			
2. Get angry or wild real quick?	Not much, sometimes, fair bit, lots of times	Depression		E. Do you get angry or wild real quick?	Not much (or never) (1), sometimes (2), fair bit (3), lots of times (4)
3. Hard to focus. Thinking all over the place.	Not much, sometimes, fair bit, lots of times	Depression			
4. Had too many bad moods?	Not really, sometimes, fair bit, lots of times	Depression			
5. Felt pretty lonely much of the time?	Not really, sometimes, fair bit, lots of times	Depression			
6. Have you felt so sad that nothing could cheer you up?	Not much, sometimes, fair bit, lots of times	Anxiety		F. Have you felt so sad that nothing could cheer you up? Not even your friends make you feel better.	Not much (or never) (1), sometimes (2), fair bit (3), lots of times (4)
7. Have you felt so worried you start to shake?	Not really, sometimes, fair bit, lots of times	Anxiety			
8. Have you felt so worried it was hard to breathe?	Not really, sometimes, fair bit, lots of times	Anxiety		D. Have you felt so worried it was hard to breathe?	Never (1), little bit (2), fair bit (3), lots (4)
9. Have you felt so worried you got really sweaty?	Not much, sometimes, fair bit, lots of times	Anxiety			
10. Have you been so worried you felt sick in the guts?	Not really, sometimes, fair bit, lots of times	Anxiety		C. Have you felt so worried your stomach (tummy) has got upset? Big worries make you sick.	Never (1), little bit (2), fair bit (3), lots (4)

11. Have you felt so worried you got dizzy?	Not really, sometimes, fair bit, lots of times	Anxiety			
12. Got angry or wild and stayed that way for a long time?	Not really, sometimes, fair bit, lots of times	Depression			
13. Felt like giving up - no point in trying?	Not really, sometimes, fair bit, lots of times	Depression			
14. Have you wished you were dead?	Not much, sometimes, fair bit, lots of times	Suicide risk			
15. Felt like hurting yourself?	Not really, sometimes, fair bit, lots of times	Suicide risk			
16. Have you felt like killing yourself?	Not much, sometimes, fair bit, lots of times	Suicide risk			
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	D. 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	E. 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	F. You are really into something (like music, cars, clothes, football, fishing, computers, etc.)? If necessary: Is there something you really like doing?		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.		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	J. You got an older (other) 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 got 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 or upset 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 in your life fairly quickly (easy). Like when you went to boarding school, or when you had a baby.		Always (1), most times (2), sometimes (3), not really (4)
[27.] People say you are really good at something		Resilience	I. People say that you are really good at something. Sports or fishing or looking after kids or something like that.		Always (1), most times (2), sometimes (3), not really (4)

[28.] Stopped doing things that used to be fun		--		A. Have you stopped liking things that used to be fun? Don't want to go fishing; don't want to hang out with your mates.	Never (1), little bit (2), fair bit (3), lots (4)
[29.] Felt like everything is hard work		--		B. Have you felt like everything is hard work (even little jobs are too much)? Felt too lazy to do anything?	Never (1), little bit (2), fair bit (3), lots (4)
[30.] Had really bad dreams		--			
[31.] Some part of your body is always hurting		--			
[32.] Been treated unfairly because you are Aboriginal		--			
[33.] Felt like you got no control		--			
[34.] You wish you were a different person		--			
[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 or upset, you're able to find something that cheers you up (makes you happy).		Always (1), most times (2), sometimes (3), not really (4)
[Preliminary 2. ***] You know a lot about your Aboriginal culture.		Resilience	H. You know a lot about your (Aboriginal/ Torres Strait Islander) family history and culture. **		Lots (1), fair bit (2), little bit (3), not much (4)
--				G. Do you do silly things without thinking that you feel shame about the next day?	Not much (or never) (1), sometimes (2), fair bit (3), lots of times (4)

The LSIC items presented here are those used in Wave 8; there is some variation in the prompt or in the clarification text used across waves. Bolded text indicates modified text included in LSIC. Blue text indicates clarifying statements used in LSIC when required. * Earlier waves included the item "You know lots about whitefella ways" (Strong Souls Item 17), which was asked of Aboriginal and/or Torres Strait Islander P1s only. This item was dropped in subsequent waves in response to RAO feedback; as such, we have not included this item in the table or in our analysis. ** Only asked of Aboriginal and/or Torres Strait Islander P1s. Non-Indigenous P1s were asked, "You understand a lot about (Aboriginal/Torres Strait Islander) history and culture". *** 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 Distress items in LSIC

The Strong Souls Distress items were collected from P1s at Waves 1-4, 6, 8 and 10. The results for the Strong Souls Distress items were similar at Waves 8 and 10, so only Wave 10 are included in the current Report. Appendix B presents the construct validity and internal consistency of the Strong Souls Distress at Wave 8.

Acceptability

Strong Souls Distress items were treated as non-response if the respondent answered don't know, or if the respondent refused to answer. The response option don't know had a greater number of non-response compared to refused. Overall, the extent of non-response was fairly low. There was a total of 113 non-responses (34 refused and 79 don't know), averaging 8.9 non-responses per 100 surveys.

The original pool of 34 Strong Souls items was developed in consultation with Aboriginal and Torres Strait Islander peoples, supporting the acceptability of the items. The item *Have you stopped liking things that used to be fun* had the highest level of non-responses, at 2.52 per 100 surveys. This item is one of the three items used in LSIC that was not included in the final Strong Souls instrument. The extent of non-response was lower, and similar to that of the final Strong Souls items, for the two other items used in LSIC but not in the final Strong Souls items, *have you felt like everything is hard work (even little jobs are too much)* (0.87 per 100 surveys) and, *do you do silly things without thinking that you feel shame about the next day* (1.65 non-response per 100 surveys).

Key points: The extent of non-response to the Strong Souls Distress items was low, indicating that these measures are acceptable for use with this sample. This likely reflects the developmental processes underlying development of the original pool of potential items for Strong Souls.⁶³

Construct validity

A PCA was conducted to explore the number of reliable components from the items, and to examine the relationship between the items and the components. An exploratory approach was employed as the current items did not align with any single Strong Souls subscale.

One item (*do you do silly things without thinking that you feel shame about the next day*) did not correlate with any items above the 0.3 threshold; one item (*do you get angry or wild real quick*) correlated above the 0.3 threshold with only two other items. Both items were retained for the PCA but it was anticipated that the former item would not load onto the component. Results of the Bartlett test of sphericity were significant (degrees of freedom = 21) and the KMO score of 0.84 indicated that correlations were adequate to proceed with PCA.

The number of components for extraction was determined by the three-criterion approach. All three approaches recommended a one-component solution be retained for the PCA.

As only one component was retained, rotation was not applicable. Only items with a component loading of >0.3 were interpreted. The final solution explained 46.5% of the variance and all items loaded onto the one component with the exception of *Do you do silly things without thinking that you feel shame about the next day* and *Do you get angry or wild real quick* (see Table 5.2). Both items had

an inter-item correlation below or close to below the 0.3 threshold, indicating that the items may not load well onto the component.

Table 5.2. Results from the PCA for the seven-item Strong Souls Distress measure reported by P1s at Wave 10

	Component
	1
1. Have you stopped liking things that used to be fun	0.400
2. Have you felt like everything is hard work (even little jobs are too much)	0.415
3. Have you felt so worried your stomach (tummy) has got upset	0.424
4. Have you felt so worried it was hard to breathe	0.409
5. Do you get angry or get wild real quick	
6. Have you felt so sad that nothing could cheer you up	0.425
7. Do you do silly things without thinking that you feel shame about the next day	

The PCA was re-run excluding Items 5 and 7. The solution explained 58.9% of the variance and all items loaded onto one component above 0.4 (see Table 5.3).

Table 5.3. Results from the PCA for the truncated five-item Strong Souls Distress measure reported by P1s at Wave 10

	Component
	1
1. Have you stopped liking things that used to be fun	0.435
2. Have you felt like everything is hard work (even little jobs are too much)	0.444
3. Have you felt so worried your stomach (tummy) has got upset	0.464
4. Have you felt so worried it was hard to breathe	0.449
6. Have you felt so sad that nothing could cheer you up	0.444

Key points: Two of the items in the Strong Souls Distress measure appear to be performing poorly, no loading onto the component: *Do you do silly things without thinking that you feel shame about the next day* and *Do you get angry or get wild real quick*. An alternate five-item Strong Souls Distress measure was created that excluded these two items.

Internal consistency

The internal consistency of the seven-and-five-item LSIC Strong Souls Distress measure was assessed using Cronbach’s alpha (see Table 5.4). Both scales had an internal consistency considered above “acceptable”, with the five-item version having had a slightly higher alpha.

Table 5.4. Cronbach’s alpha for the seven-item and five-item LSIC Strong Souls Distress measure.

Strong Souls Distress measure	α
Seven-item	0.80 ✓
Five-item*	0.82 ✓

*Excludes items 5 and 7. 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 both the seven-item Strong Souls Distress scale and the five-item Strong Souls Distress scale.

Convergent validity

This section included only respondents where the P1 was the same individual at Wave 10 and Wave 11, and where the respondent had a non-missing total Strong Souls Distress score at Wave 10 and self-reported clinical depression and/or anxiety score at Wave 11. A total of 50 respondents who

completed the clinical depression and/or anxiety measure at Wave 11 were different from the P1 who completed the Strong Souls Distress measure at Wave 10. These 50 P1s were excluded from the analyses.

Total scores for the seven- and five-item scales were computed by summing responses across the items. Similar to the K10, respondents missing any items had a missing total score. The seven-item scale ranged from 7 to 25, and the five-item ranged from 5 to 20, with higher scores indicating higher levels of distress. The total number of respondents was 1054; of these 94.9% ($n = 1000$) had a seven-item scale score computed and 95.5% ($n = 1007$) had a five-item scale score.

Given the lack of established cut-offs for using these scales, we derived three categories for both Strong Souls Distress versions (low, moderate and high), based on the distribution of scores within the LSIC sample. Respondents in (approximately) the bottom quartile were classified as having low distress, respondents in the middle 50% were classified as having moderate distress, and respondents in the highest quartile were classified as having high distress. For the seven-item Strong Souls Distress scale, a score of 7-8 was categorised as low distress ($n = 299$; 29.9%), 9-12 as moderate ($n = 480$; 48.0%), and 13 and above as high ($n = 221$; 22.1%). For the five-item Strong Souls Distress scale, a score of 5-6 was categorised as 'low' distress ($n = 227$; 22.5%), 6-9 as 'moderate' ($n = 564$; 56.0%) and 10 and above as 'high' ($n = 216$; 21.5%).

We examined the association between distress category, according to the two Strong Souls Distress scales, and P1 self-reported clinical depression and/or anxiety at Wave 11 (Table 5.5). We found that the percentage of self-reported clinical depression and/or anxiety increased with increasing level of distress, from 1.4% to 23.5% with increasing category on the seven-item scale, and from 1.7% to 24.5% on the five-item scale.

The odds of reporting clinical depression and/or anxiety were significantly higher among participants with moderate or high (compared to low) levels of distress. For the seven-item scale, those with moderate levels of distress had >5 times the odds of having clinical depression and/or anxiety, and those with high levels had >22 times the odds, compared to those with low levels of distress. The direction and magnitude of effect was similar for the five-item scale. We note that the confidence intervals around these estimates are quite wide. However, despite this and the use of arbitrary cut-off points, we did observe significant associations, in the hypothesised direction, between Strong Souls Distress scores and the mental health outcome.

Table 5.5. Percentage and Odds Ratio of P1s self-reported clinical depression and/or anxiety at Wave 11, by level of distress at Wave 10

Level of distress	No clinical depression and/or anxiety % (n)	Clinical depression and/or anxiety % (n)	Odds Ratio of having clinical depression and/or anxiety (95%CI)
Seven-item Strong Souls Distress scale			
Low (n=294)	98.6 (290)	1.4 (4)	1 (ref)
Moderate (n=480)	92.5 (444)	7.5 (36)	5.88 (2.07, 16.69)
High (n=217)	76.5 (166)	23.5 (51)	22.27 (7.91, 62.73)
Five-item Strong Souls Distress scale			
Low (n=222)	98.7 (219)	1.7 (3)	1 (ref)
Moderate (n=564)	93.3 (526)	6.7 (38)	5.27 (1.61, 17.26)
High (n=212)	75.5 (160)	24.5 (52)	23.73 (7.28, 77.32)

Key points: There was a strong association between level of distress and P1 mental health outcomes, according to both the seven-item and the five-item Strong Souls Distress scale, noting the wide confidence intervals.

Correlation between Strong Souls Distress and Kessler scores

We examined the correlation between the total Kessler scores and the Strong Souls Distress total scores for all P1s at Wave 10. There was a strong correlation for each combination of Kessler version and Strong Souls Distress version, suggesting there is a strong relationship between scores on these two measures (Table 5.6). This would be expected, given that both measures are intended to capture (psychological) distress, and are intended to capture symptoms of anxiety and depression. We also compared the percentage of variance shared across versions of the Kessler and Strong Souls Distress; the percentage of variance explained did not vary materially between versions (56.3% to 60.8%).

We did not examine categories of Strong Souls Distress scores in relation to categories of Kessler scores, given the arbitrary nature of the Strong Souls Distress categorisation used in this analysis.

Table 5.6. Pearson r correlation between the Kessler versions and Strong Souls Distress at Wave 10

	Strong Souls Distress seven-item scale		Strong Souls Distress five-item scale	
	Correlation (<i>r</i>)	% of variance shared	Correlation (<i>r</i>)	% of variance shared
K10	0.78	60.84	0.77	59.29
K5	0.77	59.29	0.75	56.25

The lack of alignment between the Strong Souls Distress items used in LSIC and the final Strong Souls Index makes use and interpretation of the LSIC measures difficult. There are no scoring guidelines for the final Strong Souls Index, or for the items used in LSIC; it is therefore difficult to meaningfully compare this measure to other measures of distress, such as the K10 or the K5.

The shift from employing the Strong Souls Distress items to using the Kessler items (starting in Wave 10) will pose a challenge for researchers who are aiming to examine P1 distress over time. Further research could be conducted to understand how Strong Souls Distress scores relate to Kessler scores; for example, scale equating of Strong Souls Distress and Kessler scores could be conducted to facilitate longitudinal analysis of P1 distress over time.

Key points: There was a strong correlation between Strong Souls Distress scores, as measured by the seven-item or five-item scale, and Kessler scores, as measured by the K10 or the K5. The magnitude and direction of correlation, and the percentage of variation shared, did not materially differ by the version of the measure used.

An alternate five-item version of the Strong Souls Distress scale was examined, given that two of the seven Strong Souls Distress items did not load on to the single component solution. However, we did not find any evidence that the five-item version performed better than the full seven-item version. Therefore, future analysis could utilise either the five-item version or the seven-item version, noting that neither aligns with the final Strong Souls items about anxiety and depression (6 and 8 items, respectively).⁶³

Results: assessment of Strong Souls Strengths items in LSIC

The Strong Souls Strengths items were collected from all P1s at Waves 1, 4 and 8. This analysis focuses on Wave 8, as it was the most recently collected (2015).

Acceptability

Items were treated as non-response if the respondent answered don't know or if the respondent refused to answer. In Wave 8 overall, the extent of non-response was fairly low, indicating acceptability of the Strong Souls: Strengths items. This is consistent with findings from process of developing and validating the Strong Souls Index, where overall, there was a low percentage of missing data.

The response option refused had a higher number of non-responses compared to don't know. There was a total of 78 non-responses (44 refused and 34 don't know), averaging 6.2 non-responses per 100 surveys. The item with the greatest number of non-response was *You are a good (son or daughter) to your family*, averaging 2.5 item non-responses per 100 surveys ($n = 18$ refuse and $n = 14$ don't know).

In the Strong Souls testing, the only item with a high level of non-response was *you get used to big changes in your life fairly quickly*; in LSIC, P1s non-response was uncommon for this item, averaging 0.3 item non-responses per 100 surveys.⁶³ In contrast *you are a good (son or daughter) to your family* had a low non-response for LSIC K-cohort children (average 1 non-response per 100 surveys) but greater non-response rate in LSIC P1s. It may be parents and/or family of P1s had passed away; this question may not be appropriate for the age of the respondent. Further exploration of this item may be warranted, in particular the appropriateness of asking this item to an adult respondent.

Construct validity

All available responses from P1s were included in the construct validity and internal consistency analyses. PCA was employed to explore the number of reliable components from the items, and to examine the relationship between the items and the components. An exploratory approach was employed as the current items did not fully align with the Resilience subscale in the Strong Souls Index.

Two items, *You are a good (son or daughter) to your family* and *You know a lot about your (Aboriginal/Torres Strait Islander) family history and culture*, did not correlate with any other items above the 0.3 threshold. Therefore, these items were excluded from subsequent analyses. Overall, the inter-item correlations were low; of the 45 possible inter-item correlations, only 7 were above the 0.3 threshold. However, results of the Bartlett test of sphericity were significant (degrees of freedom = 45) and the KMO score of 0.85 indicated that correlations were adequate to proceed with PCA.

The number of components for extraction was determined by the three-criterion approach. The three approaches revealed two to three components to be retained for the PCA. To enhance the interpretation of the PCA output, an orthogonal (varimax) and oblique (oblimin) rotated solutions 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.³⁴

The two component solution explained 44.9% of the variance (Table 5.7). Two items *You get used to big changes in your life fairly quickly* and *You got lots of friends*, had a component loading below 0.3. The items that loaded above 0.3 on Component 1 appeared to relate with social support. A clear underlying theme was not identified by the authors for Component 2.

Table 5.7. Results from the two-component PCA for the Strong Souls Resilience measure for P1s at Wave 8

	Component	
	1	2
1. When you get sad or upset, you're able to find something that cheers you up	0.365	
2. You have a strong family who help each other	0.488	
4. You know someone who is a really good person	0.311	
10. You got another (other) person looking out for you	0.482	
12. When you're sad or upset you have a person you can talk to	0.443	
5. You laugh and make jokes a lot		0.447
6. You are really into something (like music, cars, clothes, football, fishing, computers etc.)?		0.562
9. People say that you are really good at something		0.550
3. You get used to big changes in your life fairly quickly		
11. You got lots of friends		

The three component solution explained 54.6% of the variance (Table 5.8). Similar to the two-component solution, the items that loaded above 0.3 on Component 1 related to social support, and the items that loaded onto Component 2 were less defined. The item *you got a lot of friends* cross-loaded onto Components 1 and 2. The two items that loaded on Component 3 were both excluded from the final Strong Souls Index.⁶³

Table 5.8. Results from the three-component PCA for the Strong Souls Resilience measure for P1s at Wave 8

	Component		
	1	2	3
2. You have a strong family who help each other	0.464		
4. You know someone who is a really good person	0.361		
10. You got an older (other) person looking out for you	0.604		
12. When you're sad or upset you have a person you can talk to	0.367		
11. You got lots of friends*	0.368	0.360	
5. You laugh and make jokes a lot		0.414	
6. You are really into something (like music, cars, clothes, football, fishing, computers etc.)?		0.569	
9. People say that you are really good at something		0.562	
3. You get used to big changes in your life fairly quickly			0.614
1. When you get sad or upset, you're able to find something that cheers you up			0.654

Item cross-loaded onto two components above the 0.3 threshold.

There was some level of agreement between the two- and three-component solutions. Three items (*when you get sad or upset, you're able to find something that cheers you up, you get used to big changes in your life fairly quickly, and you got lots of friends*) appeared to load poorly in both solutions; the first two were not retained in the final Strong Souls Index scale. Four items appeared to capture social support (*you have a strong family who help each other, you know someone who is a really good person, you got an older (other) person looking out for you, and when you're sad or upset you have a person you can talk to*). The final three items appeared to measure a single construct, but it is unclear exactly what construct is being captured (*you laugh and make jokes a lot, you are really into something (like music, cars, clothes, football, fishing, computers etc.)?, and people say that you are really good at something*).

The eight items from the Resilience subscale of the Strong Souls Index (representing eight of the total nine items) were tested. The parallel analysis and scree-plot indicated a one-component solution, while the Kaiser's criterion indicated a two-component solution. The two-component solution was assessed but *you know someone who is a really good person* and *you are a good (son or daughter) to your family* did not load above the 0.3 threshold. In addition, the authors could not identify an underlying theme for each component based on the items that loaded at or above 0.3. Therefore the results are not presented here.

The one-component solution explained 35.5% of the variance. All items loaded above the 0.3 threshold with the exception of *you are a good (son or daughter) to your family* which had a loading of 0.294. Because all eight items were part of the Resilience subscale of the Strong Souls Index, all eight items were retained and used in the following analyses.

Table 5.9. PCA results for the eight-item resilience measure for all responding P1s, at Wave 8

	Component
	1
2. You have a strong family who help each other	0.371
4. You know someone who is a really good person	0.368
5. You laugh and make jokes a lot	0.372
6. You are really into something (like music, cars, clothes, football, fishing, computers etc.)?	0.330
7. You are a good (son or daughter) to your family	
10. You got an older (other) person looking out for you	0.316
11. You got lots of friends	0.395
12. When you're sad or upset you have a person you can talk to	0.371

Internal consistency

The eight-item Strong Souls Resilience measure had an internal consistency score of 0.73, indicative of an 'acceptable' standard.

Convergent validity

To test convergent validity of the eight-item resilience measure, it was examined in relation to self-reported clinical depression and/or anxiety at Wave 9. A total of 54 respondents who completed the clinical depression and/or anxiety measure at Wave 9 were different from the P1 who completed the Strong Souls Resilience measure at Wave 8. These 54 P1s were excluded from the analyses.

Scores were summed across the eight items. Respondents missing any items had a missing total score. The total score ranged from 8 to 32, with higher scores indicating greater resilience. There were 1064 total respondents; of these 96.4% ($n = 1026$) had total scale score computed.

Given the lack of established cut-offs for using the Strong Souls Index, or this subset of items, we derived three categories for the eight-item resilience measure (low, moderate and high), based on the distribution of scores within the LSIC sample. Respondents in (approximately) the bottom quartile were classified as having low resilience, respondents in the middle 50% were classified as having moderate resilience, and respondents in the highest quartile were classified as having high resilience. Because resilience is a feature of positive social and emotional wellbeing, we hypothesised that higher levels of resilience would be associated with a reduced odds of mental health conditions. For the Strong Souls Resilience scale, a score of 8-25 was categorised as low resilience ($n = 250$; 24.3%), 26-30 as moderate ($n = 519$; 50.5%), and 31 and above as high ($n = 259$; 25.2%).

The odds of reporting clinical depression and/or anxiety were significantly lower among participants with moderate or high (compared to low) levels of resilience. Compared to those with low levels of resilience, the respondents with moderate levels of resilience had 61% lower odds of having clinical depression and/or anxiety, and those with high levels of resilience had 81% lower odds. This supported our hypothesis that resilience may serve as a protective factor against clinical depression and/or anxiety (12 months later).

Table 5.10. Percentage and Odds Ratio of same P1 self-reported clinical depression and/or anxiety at Wave 9, by eight-item resilience measure category at Wave 8

Level of resilience	No clinical depression and/or anxiety % (n)	Clinical depression and/or anxiety % (n)	Odds Ratio of having clinical depression and/or anxiety (95%CI)
Strong Souls Resilience scale			
Low (n=250)	82.8 (207)	17.2 (43)	1 (ref)
Moderate (n=518)	92.5 (479)	7.5 (439)	0.39 (0.25, 0.62)
High (n=258)	96.1 (248)	3.9 (10)	0.19 (0.10, 0.40)

Discriminant validity

To test discriminant validity of the eight-item resilience measure, it was examined in relation to Strong Soul: Distress (seven-item measure) scores at Wave 8. This included all responding P1s at Wave 8.

There was a moderate negative correlation ($r = -0.35$) between the measures, providing support of discriminant validity. Thomas et al. (2010) reported a significant negative correlation between the Strong Souls Anxiety and Resilience measures ($r = -0.22$) and Strong Souls Depression and Resilience measures ($r = -0.21$).

Recommendations

Recommendations for data users

1. We recommend caution in analysing and interpreting the Strong Souls: Distress items used in LSIC. Given the lack of alignment between the measures used in LSIC and the final Strong Souls Index, the LSIC items should not be assumed to be equivalent to the mental health subscale of the Strong Souls Index. The use of item(s) from this measure and categorisation of scores would depend upon the research question.
2. For measurement of resilience, we recommend use of the eight-item resilience measure (rather than using all 12 Strong Souls: Strengths items). This eight-item measure comprises 8 of the 9 items in the Resilience subscale of the final Strong Souls Index. The remaining four Strong Souls: Strengths items could be used individually or as a group.
3. 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. 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. Glascoe FP. Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews* 2005; **11**(3): 173-9.
41. 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.

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. Andrews G, Slade T. Interpreting scores on the Kessler psychological distress scale (K10). *Australian and New Zealand journal of public health* 2001; **25**(6): 494-7.
44. Slade T, Grove R, Burgess P. Kessler psychological distress scale: normative data from the 2007 Australian National Survey of Mental Health and Wellbeing. *Australian and New Zealand Journal of Psychiatry* 2011; **45**(4): 308-16.
45. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological medicine* 2002; **32**(6): 959-76.
46. McNamara BJ, Banks E, Gubhaju L, et al. Measuring psychological distress in older Aboriginal and Torres Strait Islanders Australians: a comparison of the K-10 and K-5. *Australian and New Zealand Public Health* 2014; **38**(6): 567-73.
47. Australian Institute of Health and Welfare. Measuring the social and emotional wellbeing of Aboriginal and Torres Strait Islander peoples. Canberra: Australian Institute of Health Welfare, 2009.
48. Australian Bureau of Statistics. Information Paper: Use of the Kessler Psychological Distress Scale in ABS Health Surveys, Australia, 2007-08- Inclusion of the K5 in the ABS Surveys. Cat. No. 4817.0.55.001. In: Australian Bureau of Statistics, editor. Canberra: Commonwealth of Australia; 2012.
49. Lace JW, Greif TR, McGrath A, et al. Investigating the factor structure of the K10 and identifying cutoff scores denoting nonspecific psychological distress and need for treatment. *Mental Health & Prevention* 2019; **13**: 100-6.
50. 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.
51. Mitchell CM, Beals J. The utility of the Kessler Screening Scale for Psychological Distress (K6) in two American Indian communities. *Psychological assessment* 2011; **23**(3): 752.
52. Batterham PJ, Sunderland M, Slade T, Cate AL, Carragher N. Assessing distress in the community: psychometric properties and crosswalk comparison of eight measures of psychological distress. *Psychol Med* 2018; **48**(8): 1316-24.
53. Batterham PJ, Sunderland M, Carragher N, Cate AL, Mackinnon AJ, Slade T. The Distress Questionnaire-5: Population screener for psychological distress was more accurate than the K6/K10. *Journal of clinical epidemiology* 2016; **71**: 35-42.
54. Isaacs AN, Pyett P, Oakley-Browne MA, Gruis H, Waples-Crowe P. Barriers and facilitators to the utilization of adult mental health services by Australia's Indigenous people: Seeking a way forward. *International Journal of Mental Health Nursing* 2010; **19**(2): 75-82.
55. Kanowski LG, Jorm AF, Hart LM. A mental health first aid training program for Australian Aboriginal and Torres Strait Islander peoples: description and initial evaluation. *International Journal of Mental Health Systems* 2009; **3**(1): 10.
56. Australian Bureau of Statistics. Data Item List: National Aboriginal and Torres Strait Islander Social Survey, 2008. Cat. No. 4714.0_12_ Health. In: Australian Bureau of Statistics, editor. Canberra: Commonwealth of Australia; 2010.
57. Australian Bureau of Statistics. 4720.0 - National Aboriginal and Torres Strait Islander Social Survey: User Guide, 2014-15 In: Australian Bureau of Statistics, editor. Canberra: Commonwealth of Australia; 2016.
58. Gubhaju L, Banks E, Ward J, et al. 'Next Generation Youth Well-being Study:' understanding the health and social well-being trajectories of Australian Aboriginal adolescents aged 10-24 years: study protocol. *BMJ open* 2019; **9**(3): e028734.

59. Smout MF. The factor structure and predictive validity of the Kessler Psychological Distress Scale (K10) in children and adolescents. *Australian Psychologist* 2019; **54**(2): 102-13.
60. 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.
61. Prince M. Measurement validity in cross-cultural comparative research. *Epidemiology and Psychiatric Sciences* 2008; **17**(3): 211-20.
62. Sørliie 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.
63. 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.
64. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 1999; **6**(1): 1-55.

Appendices

Appendix A. K10 and K5 Confirmatory Factor Analysis

We assessed the single factor structure of the K10 ($N=1218$) and K5 ($N=1228$) using a Confirmatory Factor Analysis.

Four fit indexes were used to assess the fit between the K10 and K5 CFA models and the data: root mean square error of approximation (RMSEA), root mean squared residual (SRMR), comparative fit index (CFI) and Tucker-Lewis Index (TLI). A cut-off close to or less than .06 for RMSEA and .08 for SRMR, and a cut-off close to or greater than .95 for the CFI and TLI was used as a measure of adequate fit.⁶⁴ However, these cut-offs are rules of thumb and specific cut-offs for each index may differ slightly based on various conditions.⁶⁴

The chi-squared statistic was not used as a measure of fit because the null hypothesis is often rejected with a large sample size. This indicates the data is statistically unsuitable for factor analysis. As the current sample was considered large, it was expected this would result in a significant chi-squared.

The modification index was inspected to identify possible improvements to the model. The modification index provides an estimate how the model might fit better if we estimated an additional parameter, such as correlated error terms.

K10

The 10 items loaded significantly onto a single dimension with the standardised loadings ranging from 0.63-0.83. The fit of the model appeared to be unsatisfactory: RMSEA=0.13, SRMR=0.048, CFI=0.91, TLI=0.88.

The modification index indicated model improvement if two pairs of error terms were correlated (*Nervous* and *So nervous that nothing could calm you down*) as well (*Depressed* and *That everything was an effort*). These correlations also make sense conceptually, with the first two items related to anxiety and the latter two related to depression.

The K10 model was rerun with the error terms correlated. The 10 items loaded significantly onto a single dimension and the standardised loadings ranged from 0.63-0.85. The fit of the model was improved but the RMSEA and TLI were below the suggested cut-off: RMSEA=0.10, SRMR=0.041, CFI=0.95, TLI=0.93.

K5

The 5 items loaded significantly onto a single dimension and the standardised loadings ranged from 0.72-0.80. The fit of the model indicated potential for improvement: RMSEA=0.15, SRMR=0.037, CFI=0.95, TLI=0.91.

The modification index indicated a better fitting model if two depression items were correlated (*Depressed* and *That everything was an effort*).

The K5 model was rerun with the error terms correlated. The 5 items loaded significantly onto a single dimension and the standardised loadings ranged from 0.67-0.82. The fit of the model was improved with the SRMR, CFI and TLI indicating adequate model fit with the data: RMSEA=0.08, SRMR=0.018, CFI=0.99, TLI=0.97.

Appendix B. Strong Souls Distress at Wave 8

Construct validity

As per Strong Souls Distress at Wave 10, a PCA was conducted to explore the number of reliable components from the items, and to examine the relationship between the items and the components.

An alternate five-item version of the Strong Souls Distress scale was examined at Wave 10, but we did not find any evidence that the five-item version performed better than the full seven-item version. Therefore all items were retained for the current analyses and the seven-item Strong Souls Distress instrument at Wave 8 was employed to test the discriminant validity of the Strong Souls Resilience measure.

Two items had a low number of correlations above the 0.3 threshold *Do you do silly things without thinking that you feel shame about the next day* and *Do you get angry or wild real quick*. Both items were retained for the PCA but it was anticipated *Do you do silly things without thinking that you feel shame about the next day* would not load onto the component. Results of the Bartlett test of sphericity were significant (degrees of freedom = 21) and the KMO score of 0.84 indicated that correlations were adequate to proceed with PCA.

The number of components for extraction was determined by the three-criterion approach (see Chapter 3 Methods). All three approaches revealed a one-component solution be retained for the PCA.

As only one component was retained, rotation was not applicable. Only items with a component loading of >0.3 were interpreted. The final solution explained 45.5% of the variance and all items loaded onto the one component with the exception of *Do you do silly things without thinking that you feel shame about the next day* (see Table A1). This item had an inter-item correlation below or close to below the 0.3 threshold, indicating the item may not load well onto the component.

Table A1. Results from the PCA for the seven-item Strong Souls Distress measure reported by P1s at Wave 8

	Component
	1
1. Have you stopped liking things that used to be fun	0.366
2. Have you felt like everything is hard work (even little jobs are too much)	0.424
3. Have you felt so worried your stomach (tummy) has got upset	0.437
4. Have you felt so worried it was hard to breathe	0.411
5. Do you get angry or get wild real quick	0.302
6. Have you felt so sad that nothing could cheer you up	0.412
7. Do you do silly things without thinking that you feel shame about the next day	

These results were similar to Wave 10 with the exception of *Do you get angry or get wild real quick* which had a component loading of above 0.3 at Wave 8 but not Wave 10. The responding P1 may change at each Wave, so the sample of P1s at Waves 8 and 10 are slightly different. It is unknown the percentage of matching P1s at Waves 8 and 10 as this information is not available in the current data.

Internal consistency

The seven-item Strong Souls Distress measure had an internal consistency score of 0.80, indicative of an 'acceptable' standard. The internal consistency score at Wave 8 and Wave 10 were the same.