


# The Brief Evaluation of Adolescents and Children Online (BEACON): Psychometric development of a mental health screening measure for school students

Ronald M. Rapee<sup>1</sup>  | Rebecca Kuhnert<sup>1</sup> | Susan H. Spence<sup>2</sup> | Ian Bowsher<sup>3</sup> | John Burns<sup>1</sup> | Jennifer Coen<sup>4</sup> | Julie Dixon<sup>5</sup> | Pauline Kotselas<sup>6</sup> | Catherine Lourey<sup>5</sup> | Lauren F. McLellan<sup>1</sup> | Cathrine Mihalopoulos<sup>7</sup> | Lorna Peters<sup>1</sup> | Traci Prendergast<sup>6</sup> | Tiffany Roos<sup>8</sup> | Danielle Thomas<sup>9</sup> | Viviana Wuthrich<sup>1</sup>

<sup>1</sup>Centre for Emotional Health, School of Psychological Science, Macquarie University, Sydney, New South Wales, Australia

<sup>2</sup>Australian Institute of Suicide Research and Prevention and School of Applied Psychology, Griffith University, Mount Gravatt, Queensland, Australia

<sup>3</sup>Sydney Secondary College, Glebe, New South Wales, Australia

<sup>4</sup>Wellbeing and CVE, Catholic Schools NSW, Sydney, New South Wales, Australia

<sup>5</sup>The Mental Health Commission of NSW, Gladesville, New South Wales, Australia

<sup>6</sup>Psychology and Wellbeing Services, NSW Department of Education, Sydney, New South Wales, Australia

<sup>7</sup>Monash University Health Economics Group, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

<sup>8</sup>The Association of Independent Schools of NSW, Sydney, New South Wales, Australia

## Abstract

This paper describes the development and psychometric evaluation of a brief self-report measure (BEACON) to inform universal mental health screening in schools. Items assess symptoms and impairment associated with anxiety and attention/hyperactivity problems (grades 4–11) as well as depression and eating difficulties (grades 6–11), with optional items for suicidality and self-harm (grades 7–11). Initial item examination based on Item Response Theory (IRT) and classical test theory involved 3844 students in grades 4 through 11 (Study 1) and identified 18 items for grades 4–5 and 31 items for grades 6–11 that fulfilled pre-set criteria. Study 2 extended testing with 10,479 students in grades 4–11 and added an additional four items assessing impairment associated with eating difficulties for older students (grades 6–11) creating a total of 35 items for grades 6–11. All items, for both grade-level versions,

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. Journal of Clinical Psychology published by Wiley Periodicals LLC.

<sup>9</sup>Perinatal, Child and Youth, Mental Health Branch, NSW Ministry of Health, St Leonards, New South Wales, Australia

#### Correspondence

Ronald M. Rapee, School of Psychological Science, Macquarie University, Sydney, NSW 2109, Australia.

Email: [Ron.Rapee@mq.edu.au](mailto:Ron.Rapee@mq.edu.au)

#### Funding information

Bupa Health Foundation

met the pre-set criteria for IRT and classical test theory analysis supporting their strength in the measurement of the dimensions of concern. The measure showed good reliability (subscales alphas .87 to .95). Validity was also demonstrated against standard symptom measures, school grades, school absenteeism, and help-seeking. The BEA-CON appears to be a psychometrically sound measure to use in the first stage of school-based screening for mental health problems.

#### KEYWORDS

adolescents, assessment, children, mental health, psychometric properties, screening, universal screening

## 1 | INTRODUCTION

Mental disorders start early in life and affect around 13% of children and adolescents worldwide (Polanczyk et al., 2015). The early onset of mental disorder often heralds a chronic course of ongoing mental health difficulties combined with cascading life impairment (Herman et al., 2021). Critically, initial help for a mental disorder often does not occur until well into adulthood and most children and adolescents with mental disorders, especially internalizing disorders, do not get appropriate help (Ford et al., 2007; Merikangas et al., 2011; Reardon et al., 2020). Increasing the identification of mental disorders earlier in life and using that identification to guide children and adolescents to appropriate help and early intervention would facilitate prevention of a lifetime of problems and avert serious personal and economic burden.

To achieve this goal, accurate identification of children and adolescents experiencing mental health problems as early as possible is a critical first step. Universal mental health screening has been proposed as a key method to improve early identification (Kamphaus et al., 2014). Universal mental health screening involves delivering assessment to broad cross-sections of children and adolescents (universal) to proactively identify those who are suffering from, or at high risk for, mental disorder. This screening is most commonly delivered through schools due to their unique access to almost all children and adolescents in a society along with their organized logistic structures (Levitt et al., 2007; Soneson et al., 2018). Across recent years there has been a growing recognition that, alongside their educational mandates, schools play a critical role in fostering the mental health of their students (Anderson et al., 2019; Productivity Commission, 2020).

Despite the face validity of school-based, universal mental health screening, it is used surprisingly rarely. For example, surveys in the United States have reported only 2%–13% of schools engage in regular screening for mental health of their students (Bruhn et al., 2014; Romer & McIntosh, 2005). A similar lack of screening has been reported in the United Kingdom (Soneson et al., 2020). In Australia, a recent survey reported that only 15% of school psychologists had ever delivered any screening and less than 50% of these reported screening more than twice (Burns & Rapee, 2022). Even among those that do deliver some form of screening, many fail to use psychometrically developed measures (Bruhn et al., 2014; Burns & Rapee, 2022). A number of barriers are described by school mental health personnel to the delivery of regular mental health screening. Among the most common are: lack of knowledge about which measures to use and how to go about implementation; lack of devoted funds within schools along with the potential high costs of commercial measures; concern about a lack of acceptance of screening from school administration, parents, and students; uncertainty about how to score tests or interpret data (Bruhn et al., 2014; Burns & Rapee, 2022). Therefore, any screening system must address these barriers.

Although screening systems have varied slightly in their structure and processes, some form of questionnaire-based instrument(s) is included as a first step in the majority of systems. As in the broader children and adolescents mental health assessment literature, questionnaires can be targeted toward capturing various informants, each with their own perspectives and biases. While some evidence suggests that combining information from multiple informants can lead to more accurate identification (Goodman et al., 2003; Scott et al., 2009), perhaps surprisingly, some studies have failed to support this contention (Dowdy et al., 2016; Kilgus et al., 2018). From a pragmatic perspective, schools are most easily able to capture student self-reports about their mental health. Supplementing this information with either teacher or caregiver reports is time consuming, often impractical, and raises several barriers (Deighton et al., 2013). Further, there is a common perception that young people will be the most accurate reporters of their own internal symptoms, including many aspects of anxiety, depression, eating difficulties, suicidal and self-harm tendencies, and attention difficulties (Levitt et al., 2007; Renshaw & Cook, 2018). In contrast, parents and teachers likely provide critical information about more overt problems such as oppositional and conduct problems (Phares & Danforth, 1994).

One of the critical issues in developing or identifying strong measures for universal mental health screening is to clarify their primary purpose. Some universal screening programs aim to identify risk or protective factors, which are critically important for the implementation of school-wide prevention (Burns & Rapee, 2016; Herman et al., 2021). However, the majority aim to identify individual students who manifest high levels of symptoms of psychopathology and hence who may have or be at increased risk for mental disorders and who may not be currently accessing mental health support or intervention (Scott et al., 2009). For this purpose, direct assessment of problematic behavior, symptoms, or mental disorder is likely to hold the most promise. When identifying mental disorder, it is critical to recognize that diagnosis of a disorder and requirement of help-seeking requires more than simply expressed behaviors or symptoms. Extensive work points to the critical role of functional impairment in both defining a mental disorder and in determining need for help (Angold et al., 1999; Rapee et al., 2012). In fact, families commonly identify the impact of symptoms as more important to their treatment goals than the symptoms themselves (Creswell et al., 2021; Dickson et al., 2022) and account for more variance than symptoms in diagnostic assignment (Evans et al., 2017). Further, including impairment ratings within mental health assessment can dramatically reduce identified populations over pure symptom scores and bring them closer to accepted prevalence (Bird et al., 1990; Costello, 1996). Therefore, universal mental health screening measures would benefit from including some form of impairment rating into their scoring.

A number of self-report questionnaires have been used in universal, mental health screening, many of which have been taken from general clinical or epidemiological survey tools, and only a handful of which have been specifically developed with universal screening in mind. Most measures show clear factor structure, strong reliability, and good convergent validity (agreement with other, similar measures) (Jenkins et al., 2014; Levitt et al., 2007). Unfortunately, some of the older and better-performing measures are lengthy and costly, reducing their practicability in a school setting (Herman et al., 2021). Some more recent measures have been developed that are brief and free, but some focus on only a specific form of disorder (Renshaw & Cook, 2018; Spence & Rapee, 2022; Williamson et al., 2022), or are applicable to a limited age range (Renshaw & Cook, 2018; Williamson et al., 2022). None currently incorporate items assessing functional impairment within their core measure, although the Strengths and Difficulties Questionnaire (Goodman et al., 2003) does include additional impairment items as a supplement.

Perhaps the measure that most extensively addresses the identified barriers to mental health screening is *Me and My School* (M&MS) (Deighton et al., 2013). M&MS contains 16 items and provides a broad measure of general mental health comprising two subscales of emotional and behavioral difficulties. The measure has been validated in children aged from 8 to 15 years and shows solid psychometric properties on all standard metrics. Despite its many strengths, M&MS still has some limitations. First, although the primary purpose of screening is to provide an initial identification of young people who may have some type of mental disorder, more fine-grained guidance as to the specific nature of the potential disorder could be valuable for some purposes. For example, within internalizing

difficulties, distinguishing anxiety, depression, and eating difficulties may hold implications for schools wishing to follow screening with specific early intervention programs. Second, as noted earlier, assessment that only includes symptoms of a problem without considering functional impairment, can be overinclusive.

The primary aim of the current study was to develop a universal mental health screening measure for children and adolescents that was brief, covered the most common, self-identified, mental disorders, and could be made freely available. Although questionnaires can never replace clinical assessment in the diagnosis of mental disorder, the primary purpose of the measure was to identify students who are showing indicators of an existing mental disorder such that they may be considered to have “possible” mental disorder. In turn, this should identify students who would benefit from further, more detailed, assessment. Given the complications of collecting information from parents or teachers, we decided to focus on creating a self-report measure for students. In turn, this format led us to restrict development to grades four and above, rather than the early elementary grades due to a concern for the validity of self-report in younger children. Although the primary aim was to develop a measure that resulted in a broad identification of mental health problems, we also sought to incorporate subscales reflecting several specific forms of mental disorder to facilitate guidance toward help-seeking. Based on suggestions that children and adolescents are better at reporting internal experiences and are less reliable at reporting externalizing behaviors, we focused the measure on difficulties relating to anxiety, depression, and eating difficulties, and also included items assessing attention/hyperactivity problems, which even though they are usually referred to as “externalizing,” have a significant, experiential component. In response to requests from schools, we also added additional items that could be optionally incorporated by schools, assessing self-harm and suicidality. Most importantly, we ensured that items covered not only symptoms but also functional impairment. We named the measure the Brief Evaluation of Adolescents and Children Online (BEACON).

We report this paper as two studies. The first study describes the development and initial psychometric evaluation of the BEACON and the second study provides replication and additional analyses within a larger, independent sample.

## 2 | STUDY 1

### 2.1 | Method

#### 2.1.1 | Participants

Participants for Study 1 included 3844 students (54.6% male, 42.7% female, 1.7% other;  $M_{\text{age}} = 11.50$ ;  $SD = 3.01$ ) in grades 4 through 11 from 30 schools in the state of New South Wales, Australia. Schools volunteered for participation in the study following information in schools' newsletters and were therefore not perfectly representative of the population. However, volunteer schools were recruited across the three major education sectors in Australia: public ( $n = 23$ ), independent ( $n = 6$ ), and Catholic ( $n = 1$ ) systems. They also represented metropolitan ( $n = 21$ ), regional ( $n = 8$ ), and remote ( $n = 1$ ) geographic areas (Australian Bureau of Statistics, 2016a), and postcodes that were socioeconomically above average ( $n = 11$ ), average ( $n = 10$ ), and below average ( $n = 9$ ) (Australian Bureau of Statistics, 2016b). Students self-reported their ethnic affiliation as Australian (73.0%), First Nations (4.5%), Asian (10.2%), European (3.3%), Middle Eastern (1.0%), African (1.1%), and other (6.9%).

#### 2.1.2 | Measures

##### *Demographics*

Before screening, students completed demographic questions online which included grade, date of birth, gender, ethnicity, and parent/caregiver living arrangement.

## BEACON

The BEACON was developed to assess students' self-reported mental health across the dimensions of anxiety, depression, eating difficulties, and attention/hyperactivity problems. Items were incorporated or adapted from previous psychometrically sound measures (e.g., Spence Children's Anxiety Scale, Spence, 1998; Brief Emotional Distress Scale for Youth, Spence & Rapee, 2022; Children's Anxiety Life Interference Scale, Lyneham et al., 2013) or were developed by the research team based on DSM-5 criteria.

An initial pool of 37 items for students in grades 4–5 (approx. ages 9–11 years), 49 items for students in grade 6 (approx. age 11–12 years), and 51 items for students in grades 7–11 (approx. ages 12–17 years) were selected (see Supporting Information: Tables 3–6 for the full item banks). Items assessed symptoms of mental disorder in relation to anxiety, depression, attention/hyperactivity problems, and eating disorders (referred to as eating difficulties here). Each of the first three sets of symptoms were also accompanied by six items assessing the degree of life impairment associated with each symptom domain. These impairment items assessed general distress associated with the specific symptoms in addition to their impact on peer relationships, family relationships, health, daily activities, and positive activities (e.g., "When you felt scared or worried, how hard was it for you to get on with friends?"). Because impairment associated with eating disorder symptoms is rarely assessed, we omitted impairment items for these symptoms in Study 1 but added them in Study 2. All items were rated on 5-point Likert scales ranging from 1 ("Not at all" for symptom items; "Not at all hard" for impairment items) to 5 ("A great deal" for symptom items; "Very hard" for impairment items). Because eating disorders and major depression are rare before adolescence (Rapee et al., 2019) and to minimize potential distress, students in grades 4 and 5 did not receive these items and students in grade 6 did not receive the eating difficulties items in Study 1 (they did in Study 2).

**2.1.2.0.1 | Optional items.** In addition to the core BEACON, optional additional items included suicidality (two items) and self-harm (two items) and were developed by the authors based on clinical criteria. These items were only available for students in grades 7–11. The pairs of items were sequential (e.g., "Have you ever thought of killing yourself, even if you would not really do it" followed by "How seriously did you think about killing yourself" only if they answered yes) and were not designed to reflect a dimensional construct.

## 2.1.3 | Additional measures for validation

Subgroups of participants completed several measures to assess validity of the BEACON and each of its subscales (as well as additional research not described here). Schools were randomly allocated to a given set of measures to ensure that questionnaire packs were not too large.

### *Student-reported measures*

Several widely used, children and adolescent self-reported, measures of mental health were included to assess validity of each subscale. These included: Revised Child Anxiety and Depression Scale—anxiety subscale (RCADSa) and depression subscale (RCADSd; Chorpita et al., 2000); the ADHD Rating Scale-IV-Child Version (ADHD-IV-C; DuPaul et al., 1998) (which we adapted for student self-report), the Work and Social Adjustment Scale for Youth (WSASY; De Los Reyes et al., 2019), and the Eating Disorder Examination Questionnaire (EDE-Q; Fairburn & Cooper, 1993) (grades 7–11 only). In addition to these measures of mental health, students also reported on several related constructs including peer victimization—selected items from the Peer Experiences Checklist (PECK) (Hunt et al., 2012); school connectedness, family relations, and peer relationships subscales from the RADAR (Burns & Rapee, 2016, 2019); and quality of life via the Child Health Utility instrument (CHU-9D; Stevens, 2010).

### *Parent-reported measures*

Parallel, parent-reported versions of the above measures were included: Spence Children's Anxiety Scale, parent report (SCASp; Nauta et al., 2004); Short Mood and Feelings Questionnaire, parent report (SMFQ-P; Angold et al., 1995), and the ADHD Rating Scale-IV-Parent Version (ADHD-IV-P; DuPaul et al., 1998).

### *School measures*

Students whose parents provided opt-in consent also had their school contacted by the researchers to obtain the number of days absent from school over the past school term; and their school grades at the last assessment.

## 2.1.4 | Procedure

Ethical approval for data collection was obtained from the Macquarie University Human Research Ethics Committee, the NSW State Education Research Applications Process, and relevant independent school ethics processes. Within each school, students were included in BEACON screening through opt-out consent. This was accepted by the two main ethics committees based on extensive evidence that opt-in consent commonly excludes children and adolescents who are at the greatest risk for mental health concerns (Moore et al., 2015; Soneson et al., 2020). Children and adolescents were included in screening if they: (1) were enrolled in grades 4 through 11 at the time of testing; (2) were not "opted-out" by their caregiver; (3) assented themselves to the testing; and (4) were present at school on the day of testing. Collection of data for parent and school measures was done through a separate opt-in consent process, where parents who wished to participate in these components were required to provide written consent for themselves and for access to school data on behalf of their child. Due to the opt-in nature of this additional component, these data were collected on a small subsample of the larger sample. For participation in the opt-in components of the study, families were remunerated with an AUD\$60 gift voucher.

Students completed the BEACON and additional student measures during school hours on school grounds. Each student completed screening on a separate computer or device in their classroom, with a supervising teacher present for troubleshooting or to support any students if they became distressed. In the weeks following screening, an email was sent to consenting parents with the URL link to the online parent survey, which they conducted in their own time. In conjunction, schools provided data on academic performance and absenteeism for students whose parents had consented for their child's data to be released.

All study data were collected and managed using REDCap electronic data capture tools hosted at Macquarie University (Harris et al., 2009). Ethically and as part of a broader study of mental health screening, children and adolescents who scored in the "risk" range on the BEACON (top 15%–20%) were given feedback through their school contact who received sensitive instructions about providing the feedback to caregivers (details about the overall screening process and results of a randomized trial are reported in a separate paper, Rapee et al., 2024). Feedback about any young person scoring high on suicidality was provided immediately, and for those high on self-harm, within 24 h, whereas feedback about more general potential mental health status was provided within 2 weeks.

## 2.1.5 | Data analytic procedure

### *Step 1: Examining item characteristics using IRT*

The properties of each of the items from the first version of the BEACON screener were examined using Item Response Theory (IRT) analysis, conducted using Stata-17 (StataCorp, 2021). Item characteristic curves (ICCs) were generated for every item, separately for primary school students (grades 4–6) and secondary school students (grades 7 to 11). Given that ICCs need to be conducted on unidimensional constructs, the ICCs were generated

separately for each mental health dimension (i.e., anxiety, depression, eating difficulties, and attention/hyperactivity problems), (strong correlations between symptoms and impairment in preliminary analyses meant that we combined these as reflecting the same dimension). Five independent raters examined the ICCs to ensure that they met the following criteria:

- (a) clear separation of ICC response probability curves for response options (1–5) in the region of theta (the latent construct) of most interest for identifying elevated symptoms and impairment, between  $-1$  to  $+2$ , with clearly separated location points (difficulty thresholds, i.e., The theta value at which individuals show a 50% probability of selecting that particular response option).
- (b) steep slopes providing good discrimination of scores for response options across theta  $-1$  to  $+2$ . Items were discarded if a minimum of four of the five raters identified an item as performing poorly against these criteria.

Items that met initial IRT criteria were also examined separately by each individual grade to investigate age-related response patterns. ICCs for grade 4 and grade 5 performed similarly to one another and distinctly from the remaining grades; thus, a decision was made to include grade 6 students with the older cohort and to assess items separately for students in grades 4 to 5 (approx. ages 9–11 years) and for students in grades 6–11 (approx. ages 11–17 years). Although we allowed the possibility that selected items would differ between the two age groups, we biased our decisions to try and retain the same items where possible. Ultimately, the same set of items showed the strongest curves for both.

#### *Classical test theory analysis*

All classical test theory analyses were conducted using SPSS26 (IBM Corp., 2019). To be retained items were also required to meet pre-set criteria in either age group, specifically:

- (a) acceptable skewness and kurtosis: skew between  $-2$  to  $+2$  and kurtosis between  $-7$  to  $+7$  (Byrne, 2010).
- (b) item-total correlations  $>0.30$  for the items on the scale, with a total reliability of the scale at  $\alpha > .80$  (Nunnally, 1978).

## 2.2 | Results

For children in grades 4–5, 10 anxiety and 8 attention/hyperactivity symptom and impairment questions met all criteria and were retained (see Table 1). For grades 6–11, 10 anxiety, 9 depression, 8 attention/hyperactivity problems, and 4 eating difficulties items were also retained (see Table 2). As can be seen in the two tables, all items showed clear IRT curves, indicating good representation of its underlying construct for every item. Consequently, correlations between each item and its relevant subscale were strong (all  $>0.5$ ) and subscale reliabilities were therefore excellent (all  $>0.85$ ). Each subscale showed good variability and normal distribution. Validity of the BEACON was supported by significant correlations between the total and subscale scores with relevant alternate measures (see Tables 3 and 4). For example, RCADS-anxiety showed stronger correlations with the BEACON anxiety subscale than with other subscales and the ADHD child reports showed strongest correlations with BEACON attention/hyperactivity problems. Naturally, parent-report scores correlated more poorly with BEACON scores and in many cases were not significant. However, some significant, relevant relationships were demonstrated. For older (but not younger) students, the two proxy correlates of poor mental health, low school grades and school absences, correlated significantly with the total score and most of the subscales. The suicidality and self-harm items were not designed as a dimensional scale and were therefore fully retained. The complete BEACON can be freely obtained from <https://www.mq.edu.au/research/research-centres-groups-and-facilities/centres/lifespan-health-and-wellbeing/our-resources-and-training/children-and-teens>.

**TABLE 1** Study 1 results for analyses derived from item response theory and classical test theory for final items for grades 4–5.

	Good ICCs slope and spread	Mean	SD	Skew	Kurtosis	Item-total correlation
Anxiety (10 items)		20.62	7.72	0.81	0.19	$\alpha = 0.89$
Worried about things	Yes	2.49	1.01	0.25	-0.49	0.69
Felt afraid	Yes	1.95	0.97	0.84	0.19	0.67
Afraid of making a fool of myself	Yes	2.23	1.26	0.74	-0.57	0.60
Scared for no reason at all	Yes	1.75	1.09	1.40	1.03	0.61
Felt nervous	Yes	2.30	1.11	0.58	-0.36	0.64
Bothered you when you felt scared or worried	Yes	2.36	1.09	0.56	-0.45	0.71
Hard to get on with parents and family	Yes	1.91	1.04	1.10	0.57	0.59
Hard to get on with friends	Yes	1.90	1.10	1.20	0.07	0.56
Hard to look after your health	Yes	1.98	1.19	1.04	0.07	0.62
Hard to do activities you enjoy	Yes	1.73	1.11	1.53	1.44	0.54
Attention/hyperactivity problems (8 items)		16.12	6.45	0.98	0.71	$\alpha = 0.87$
Got distracted easily	Yes	2.45	1.17	0.57	-0.55	0.65
Hard to manage my time properly	Yes	2.05	1.13	0.93	-0.01	0.64
Didn't enjoy tasks where had to concentrate	Yes	2.22	1.27	0.77	-0.56	0.59
Hard to be still when needed to be	Yes	2.09	1.23	0.98	-0.08	0.64
Bother you when you couldn't focus or stay still	Yes	2.23	1.12	0.74	-0.20	0.66
Hard to look after your health	Yes	1.79	1.10	1.42	1.24	0.57
Hard to do daily activities	Yes	1.76	0.99	1.38	1.47	0.64
Hard to do activities you enjoy	Yes	1.52	0.92	1.20	0.06	0.55

Abbreviation: ICC, item characteristic curve.

## 2.3 | Discussion

Psychometric analysis provided strong support for 18 items assessing anxiety and attention/hyperactivity problems in the BEACON version for grades 4–5 ( $M_{\text{age}} = 9.82$ ; ages approximately 9–11) and 31 items assessing anxiety, depression, attention/hyperactivity problems, and eating difficulties for grades 6 and above ( $M_{\text{age}} = 12.73$ ; ages approximately 11–17). According to IRT analyses, items showed good representation of the relevant construct and clear discrimination between different individuals with different levels of each construct. The classical psychometric properties of the BEACON for both age groups were solid. Only one item in the attention/hyperactivity problems subscale for the grade 6–11 version showed a slightly higher skew than acceptable; however, given that its other properties were acceptable, we decided to retain it until replication in Study 2. Overall, the measure showed well-distributed items and excellent reliability for all subscales.

The BEACON also showed generally good construct validity. As expected (for both versions), there were moderate to strong correlations between each subscale of the BEACON and their corresponding construct as



**TABLE 2** Study 1 results for analyses derived from item response theory and classical test theory for final items for grades 6–11.

	Good ICCs slope and spread	Mean	SD	Skew	Kurtosis	Item-total correlation
Anxiety (10 items)		21.24	8.60	0.80	-0.02	$\alpha = 0.92$
Worried about things	Yes	2.75	1.11	0.19	-0.66	0.75
Felt afraid	Yes	1.95	0.97	0.84	0.10	0.72
Afraid of making a fool of myself	Yes	2.46	1.32	0.50	-0.92	0.66
Scared for no reason at all	Yes	1.81	1.16	1.33	0.73	0.72
Felt nervous	Yes	2.43	1.17	0.54	-0.54	0.72
Bothered you when you felt scared or worried	Yes	2.32	1.12	0.58	-0.52	0.75
Hard to get on with parents and family	Yes	1.91	1.08	1.07	0.32	0.67
Hard to get on with friends	Yes	1.80	1.02	1.24	0.87	0.66
Hard to look after your health	Yes	2.05	1.25	0.97	-0.20	0.72
Hard to do activities you enjoy	Yes	1.76	1.08	1.36	0.93	0.63
Depression (9 items)		17.65	8.90	1.09	0.29	$\alpha = 0.94$
Felt really sad	Yes	2.09	1.18	0.88	-0.19	0.82
Felt really alone	Yes	1.96	1.23	1.09	0.06	0.79
Felt like there is nothing to look forward to	Yes	2.00	1.23	1.02	-0.09	0.76
Didn't enjoy things anymore	Yes	1.85	1.15	1.24	0.54	0.77
Felt like crying a lot	Yes	1.98	1.32	1.11	-0.10	0.78
Bother you when you felt sad or lonely	Yes	2.25	1.20	0.68	-0.56	0.79
Hard to look after your health	Yes	1.90	1.21	1.25	0.47	0.78
Hard to do daily activities	Yes	1.89	1.14	1.19	0.42	0.77
Hard to do activities you enjoy	Yes	1.72	1.07	1.48	1.37	0.74
Attention/hyperactivity problems (8 items)		16.99	7.55	0.88	0.05	$\alpha = .90$
Got distracted easily	Yes	2.70	1.30	0.33	-1.01	0.76
Hard to manage my time properly	Yes	2.35	1.26	0.63	-0.69	0.69
Didn't enjoy tasks where I had to concentrate	Yes	2.39	1.34	0.59	-0.89	0.70
Hard to be still when needed to be	Yes	2.22	1.38	0.82	-0.67	0.72
Bother you when you couldn't focus or stay still	Yes	2.29	1.23	0.67	-0.59	0.73
Hard to look after your health	Yes	1.71	1.09	1.55	1.49	0.68
Hard to do daily activities	Yes	1.84	1.13	1.27	0.69	0.72
Hard to do activities you enjoy	Yes	1.47	0.89	2.13	4.17	0.60

TABLE 2 (Continued)

	Good ICCs slope and spread	Mean	SD	Skew	Kurtosis	Item-total correlation
Eating difficulties (4 items) (grades 7–11 only)		7.78	4.38	1.16	0.38	$\alpha = 0.88$
Watched what I ate so I wouldn't gain weight	Yes	2.05	1.23	0.95	-0.20	0.69
Terrified of gaining weight	Yes	1.96	1.29	1.15	0.06	0.82
Thought I'm fatter than other people think I am	Yes	2.08	1.37	0.96	-0.47	0.78
After eating, felt guilty and upset with myself	Yes	1.69	1.19	1.67	1.61	0.71

Abbreviation: ICC, item characteristic curve.

TABLE 3 Correlations between BEACON subscales and validator measures for grades 4–5.

	BEACON–total	BEACON–anxiety	BEACON–attention/hyperactivity problems
<b>Child report measures</b>			
BEACON–anxiety (n = 1616)	0.94**		
BEACON–attention/hyperactivity problems (n = 1608)	0.91**	0.70**	
Peer victimization (n = 1608)	0.55**	0.51**	0.50**
RADAR Child Version–School Connectedness Subscale (n = 1599)	-0.30**	-0.25**	-0.31**
RADAR Child Version–Family Relations Subscale (n = 1598)	-0.33**	-0.31**	-0.30**
RADAR Child Version–Peer Acceptance Subscale (n = 1598)	-0.43**	-0.40**	-0.40**
Child Health Utility Instrument (CHU-9D)–Total (n = 588)	0.73**	0.67**	0.69**
Revised Children's Anxiety and Depression Subscale (RCADS)–Anxiety Subscale (n = 553)	0.79**	0.74**	0.72**
ADHD-IV-Child version (ADHD-IV-C)–Total (n = 510)	0.68**	0.55**	0.73**
<b>Parent report measures</b>			
Spence Children's Anxiety Scale (SCAS) Parent Version–Total (n = 39)	0.20	0.21	0.16
ADHD-IV-Parent version (ADHD-IV-P)–total (n = 39)	0.27	0.15	0.37*
<b>School measures</b>			
Mean school grades (n = 71)	-0.12	-0.12	-0.05
Total absences from school (n = 65)	0.18	0.21	0.13

\*\*p < .01; \*p < .05.

**TABLE 4** Correlations between BEACON subscales and validator measures for grades 6–11.

	BEACON— total	BEACON— anxiety	BEACON— depression	BEACON— attention/ hyperactivity problems	BEACON —eating difficulties
<b>Child report measures</b>					
BEACON—anxiety ( <i>n</i> = 2187)	0.93**				
BEACON—depression ( <i>n</i> = 2173)	0.95**	0.86**			
BEACON— attention/hyperactivity problems ( <i>n</i> = 2155)	0.88**	0.74**	0.77**		
BEACON—Eating Difficulties ( <i>n</i> = 1350)	0.70**	0.56**	0.59**	0.53**	
Peer victimization ( <i>n</i> = 2160)	0.55**	0.49**	0.54**	0.48**	0.40**
RADAR Youth Version—School Connectedness Subscale ( <i>n</i> = 1345)	-0.51**	-0.47**	-0.50**	-0.43**	-0.36**
RADAR Youth Version—Family Relations Subscale ( <i>n</i> = 1345)	-0.48**	-0.42**	-0.47**	-0.44**	-0.35**
RADAR Youth Version—Peer Acceptance Subscale ( <i>n</i> = 1346)	-0.42**	-0.41**	-0.41**	-0.36**	-0.25**
Child Health Utility Instrument (CHU-9D)—total ( <i>n</i> = 641)	0.80**	0.73**	0.77**	0.75**	0.53**
Revised Children's Anxiety and Depression Subscale (RCADS)— Anxiety Subscale ( <i>n</i> = 605)	0.83**	0.81**	0.77**	0.72**	0.60**
Revised Children's Anxiety and Depression Subscale (RCADS)— Depression Subscale ( <i>n</i> = 620)	0.87**	0.79**	0.84**	0.81**	0.62**
ADHD Rating Scale-IV (ADHD-RS- IV) Child Version—total ( <i>n</i> = 566)	0.76**	0.67**	0.69**	0.82**	0.47**
Eating Disorder Examination Questionnaire (EDE-Q)— Total ( <i>n</i> = 53)	0.45**	0.43**	0.40**	0.39**	0.45**
Work and Social Adjustment Scale for Youth (WSASY)— total ( <i>n</i> = 309)	0.68**	0.63**	0.63**	0.65**	0.44**
<b>Parent report measures</b>					
Spence Children's Anxiety Scale (SCAS) Parent Version— total ( <i>n</i> = 45)	0.34*	0.38*	0.38*	0.12	0.22
Short Mood and Feelings Questionnaire (SMFQ)— total ( <i>n</i> = 45)	0.49**	0.46**	0.53**	0.40**	0.01
ADHD-RS-IV Parent Version— total ( <i>n</i> = 45)	0.14	0.12	0.18	0.10	0.01

TABLE 4 (Continued)

	BEACON— total	BEACON— anxiety	BEACON— depression	BEACON— attention/ hyperactivity problems	BEACON —eating difficulties
School measures					
Mean school grades ( $n = 72$ )	-0.28*	-0.19	-0.35**	-0.25*	-0.25
Total absences from school ( $n = 70$ )	0.39**	0.32**	0.40**	0.36**	0.37*

\*\* $p < .01$ ; \* $p < .05$ .

assessed by established measures self-reported by the students. The poorest relationship was shown between the eating difficulties subscale of the BEACON and the Eating Disorder Examination Questionnaire, although this was still a moderate relationship. This slightly lower correlation relative to the other subscales may reflect the small number of items and lack of items reflecting impairment in this version. Parent-reported measures of relevant constructs showed markedly lower correlations with each BEACON subscale than did child and adolescent self-reports, which is not unexpected, given the commonly reported inconsistencies between reporters (De Los Reyes & Kazdin, 2005). Further, because parents were included on an opt-in basis, the sample sizes were especially small for those analyses. Nonetheless, a few of the BEACON student-report subscales were validated by their corresponding parent report. The main exception was a lack of significant correlation between the attention/hyperactivity problems subscale of the BEACON and parent-reported ADHD symptoms for the older version of the measure. Importantly, the BEACON subscales and total score showed significant relationships with several broader validators. On student self-reports, there were significant correlations between most of the BEACON subscales and measures of family relationships, peer acceptance, school connectedness, peer victimization, and quality of life. Interestingly, even the objective and less directly related constructs, school absenteeism and school grades, showed small but significant relationships with several of the subscales, especially attention/hyperactivity problems and depression among the older students.

In sum, the data from this initial study provide confidence in the use of the BEACON as a measure to screen for mental disorder-related experiences among school students. However, replication in an additional sample is warranted. In addition, further refinement of items may be useful, and it is important to remedy the omission of items to assess life impairment associated with eating symptoms among older students. To address these aims, we conducted a second study among a larger sample of students.

### 3 | STUDY 2

The primary aim of Study 2 was to replicate and extend psychometric testing of the two, newly developed versions of the BEACON in a large, independent sample of students. We decided to test additional items to assess impairment associated with eating difficulties symptoms to balance the assessment of the constructs and we also added additional items assessing self-harm and suicidality to allow greater refinement of these measures. In addition to using IRT to evaluate item properties, classical test theory was used to evaluate the reliability and validity of each version of the BEACON. Validity was evaluated against distal markers of poor mental health such as school attendance, peer victimization, peer acceptance, family relationships, and school connectedness, in addition to a parent-report version of the BEACON items.

## 3.1 | Method

### 3.1.1 | Participants

Participants for Study 2 included 10,479 students (50.7% male, 46.8% female, 1.9% other;  $M_{\text{age}} = 12.92$ ;  $SD = 3.07$ ) who had not previously conducted screening and were used as an independent sample for further psychometric testing of short-listed items. The 56 schools included in Study 2 were recruited across the three major education sectors in Australia: public ( $n = 45$ ), independent ( $n = 9$ ), and Catholic ( $n = 2$ ) systems. They also represented metropolitan ( $n = 35$ ), regional ( $n = 18$ ), and remote ( $n = 3$ ) geographic areas (Australian Bureau of Statistics, 2016a), and postcodes that were socioeconomically above average ( $n = 18$ ), average ( $n = 18$ ), and below average ( $n = 20$ ) (Australian Bureau of Statistics, 2016b). Students self-reported their ethnic affiliation as Australian (59.8%), First Nations (3.6%), Asian (21.6%), European (3.4%), Middle Eastern (2.1%), African (1.3%), and other (8.2%).

### 3.1.2 | Measures

#### *BEACON*

Students completed a version of the core BEACON comprising the final 18 items for students in grades 4–5 and 31 items for students in grades 6–11 that had been selected following analyses in Study 1. An additional six items to assess impairment from eating difficulties symptoms for older students were also added and matched the impairment items previously developed for the other subscales. Thus, students in grades 6–11 completed 37 items in total. In addition to the core measure, schools were again permitted to include the measures of suicidality and self-harm (grades 7–11 only). In Study 2, suicidality and self-harm were augmented with an additional two (suicidality, e.g., “did you have a clear plan?”) and four (self-harm, e.g., “Do you think you are likely to hurt yourself on purpose again?”) items to refine assessment of these risks.

#### *Additional measures for validation student measures*

As in Study 1 students also completed three subscales of the RADAR (Burns & Rapee, 2016, 2019) and selected items from the PECK (Hunt et al., 2012). In addition, students were asked to report the number of services from which they had received help for mental health over approximately the last year.

#### *Parent measures*

A subset of students' parents who volunteered participation completed a parent-report version of the BEACON ( $n = 146$ ). This measure contained identical items to the student-report BEACON but was worded to describe “your child” rather than self. Parents were also asked to report the number of services from which their child had received help for mental health over the last 3 months.

#### *School measures*

School data on academic grades and absences was collected as per Study 1 among those students whose parents provided opt-in consent ( $n = 206$ ).

### 3.1.3 | Procedure

The same procedure was followed as outlined above for Study 1.

### 3.1.4 | Data analysis

The data analysis procedure for Study 2 mirrored that for Study 1 but with some additions to confirm the strength of the properties of the final items with the larger sample.

*Additional IRT criteria.* In addition to the criteria specified above, items were required to show:

- (a) a category characteristic curve (CCC) showing 4 clear crossover points in the theta region of interest ( $-3$  to  $+3$ );
- (b) a discrimination parameter  $>1.7$  (using STATA) indicating strong capacity to discriminate between participants with different levels of theta (Baker, 2001);
- (c) lack of significant differential item functioning (DIF: invariance) between genders, suggesting that males and females<sup>1</sup> at the same value of the latent construct theta respond equivalently in response choices on an item (Using Jmetrik). Items had to be classified as AA (reflecting a standardized  $p$ -DIF  $< .05$ ) according to the Cochran–Mantel–Haenszel chi-square statistic for DIF criteria for polytomous items (Meyer, 2014).

*Classical Test Theory Analyses:* These replicated analyses were conducted in Study 1.

## 3.2 | Results

### 3.2.1 | Grades 4–5

Results from the IRT and classical test theory analyses are presented in Table 5. For Grades 4–5 all 10 anxiety items and 8 attention/hyperactivity items met IRT criteria for the discrimination parameter, BCC and CCC curves, and DIFF between genders. For anxiety, the interference items “get on with friends” and “do activities you enjoy” showed some weakness in relation to discrimination between individuals with different levels of theta (anxiety) at higher levels of anxiety, but overall, these items were judged to be sufficiently strong to be retained in the subscale. For the classical test theory analyses, all items met all criteria in relation to skew, kurtosis, and item-total correlations. Both subscale totals met criteria for skew, kurtosis, and reliability. Correlations between the subscales and relevant validators among students in grades 4–5 are shown in Table 6.  $T$ -scores and percentiles are shown in Supporting Information: Table 1.

### 3.2.2 | Grades 6–11

First, the IRT results for the 6 newly developed impairment items associated with eating difficulties were evaluated for grades 6–11. Four items showed strong assessment of the core construct and good discrimination and were therefore retained (see Table 7), while the other two were rejected. Table 7 also shows that all Study 2 items for the dimensions of anxiety, depression, attention/hyperactivity problems, and eating difficulties showed strong IRT properties for grades 6–11, in line with Study 1. A minor exception was the depression item (I felt like crying a lot) that failed to show gender invariance on the DIFF test. As the item showed very strong performance on other IRT indicators it was retained within the BEACON depression subscale. For the classical test theory analyses, all items met all criteria in relation to skew, kurtosis, and item-total correlation. All subscale totals met criteria for skew, kurtosis, and reliability. Correlations between subscales and validators among students in grades 6–11 are shown in Table 8.  $T$ -scores and percentiles are shown in Supporting Information: Table 2.

TABLE 5 Study 2 results for analyses derived from item response theory and classical test theory for final items for grades 4–5.

	Results from item response theory		Results from classical test theory					Item-total correlation	
	Discrimination parameter	DIFF (gender)	4 CCC crossovers	Good ICCs slope and spread	Mean	SD	Skew		Kurtosis
Anxiety (10 items)					20.66	8.06	0.87	0.30	$\alpha = 0.90$
Worried about things	2.66	AA	Yes	Yes	2.49	1.06	0.33	-0.52	0.72
Felt afraid	2.41	AA	Yes	Yes	1.96	1.00	0.92	0.38	0.69
Afraid of making a fool of myself	1.61	AA	Yes	Yes	2.19	1.25	0.77	-0.51	0.60
Scared for no reason at all	2.26	AA	Yes	Yes	1.76	1.11	1.37	0.86	0.66
Felt nervous	2.05	AA	Yes	Yes	2.38	1.14	0.53	-0.48	0.65
Bothered you when you felt scared or worried	2.58	AA	Yes	Yes	2.34	1.11	0.53	-0.58	0.72
Hard to get on with parents and family	1.74	AA	Yes	Yes	1.91	1.11	1.17	0.53	0.64
Hard to get on with friends	1.35	AA	Yes	Yes	1.90	1.09	1.14	0.52	0.55
Hard to look after your health	1.70	AA	Yes	Yes	2.01	1.24	1.03	-0.07	0.63
Hard to do activities you enjoy	1.54	AA	Weak	Yes	1.73	1.10	1.53	1.45	0.58
Attention/hyperactivity problems (8 items)					16.57	6.84	0.91	0.31	$\alpha = 0.88$
Got distracted easily	1.94	AA	Yes	Yes	2.43	1.15	0.57	-0.51	0.66
Hard to manage my time properly	1.86	AA	Yes	Yes	2.12	1.14	0.81	-0.25	0.64
Didn't enjoy tasks where had to concentrate	1.68	AA	Yes	Yes	2.23	1.27	0.77	-0.49	0.62

TABLE 5 (Continued)

	Results from item response theory			Results from classical test theory					
	Discrimination parameter	DIFF (gender)	Good ICCs slope and spread	4 CCC crossovers	Mean	SD	Skew	Kurtosis	Item-total correlation
Hard to be still when needed to be	1.95	AA	Yes	Yes	2.08	1.26	0.97	-0.18	0.65
Bother you when you couldn't focus or stay still	2.28	AA	Yes	Yes	2.26	1.16	0.68	-0.49	0.69
Hard to look after your health	2.07	AA	Yes	Yes	1.80	1.10	1.34	0.94	0.65
Hard to do daily activities	2.03	AA	Yes	Yes	2.07	1.18	0.95	-0.09	0.66
Hard to do activities you enjoy	1.92	AA	Yes	Yes	1.61	0.99	1.73	2.39	0.60

Abbreviations: CCC, category characteristic curve; ICC, item characteristic curve.



**TABLE 6** Correlations between BEACON subscales and validator measures for grades 4–5.

	BEACON— total	BEACON— anxiety	BEACON—attention/ hyperactivity problems
Student self-report			
BEACON—anxiety ( <i>n</i> = 2602)	0.93**		
BEACON—attention/hyperactivity problems ( <i>n</i> = 2590)	0.91**	0.74**	
Peer victimization ( <i>n</i> = 1695)	0.55**	0.53**	0.49**
RADAR Child Version—School Connectedness Subscale ( <i>n</i> = 1666)	-0.35**	-0.28**	-0.31**
RADAR Child Version—Family Relations Subscale ( <i>n</i> = 1666)	-0.35**	-0.28**	-0.30**
RADAR Child Version—Peer Acceptance Subscale ( <i>n</i> = 1666)	-0.47**	-0.45**	-0.39**
Total number of services seen ( <i>n</i> = 2602)	0.17**	0.16**	0.16**
Parent report and school data			
BEACON parent version—anxiety ( <i>n</i> = 59)	0.41**	0.37**	0.36**
BEACON parent version—attention/hyperactivity problems ( <i>n</i> = 58)	0.32*	0.17	0.38**
Total number of services seen by their child ( <i>n</i> = 58)	0.19	0.16	0.13
Mean school grades ( <i>n</i> = 69)	-0.03	0.04	-0.12
Total absences from school ( <i>n</i> = 69)	0.11	0.10	0.10

\*\**p* < .01; \**p* < .05.

#### Validation of suicidality and self-harm risk scores

To provide an indication of the validity of the scores assessing suicidality and self-harm risk (among students in grades 7 and above, where schools opted in), we compared mean scores for those scoring “at-risk” versus the rest of the sample on the other BEACON subscales as well as the total number of services and mean school grades. At-risk students were those for whom an alert had been sent to the school. Results are shown in Table 9 and as can be seen all variables showed significant differences with large effects.

## 4 | DISCUSSION

This study presents the psychometric properties of a newly developed measure (BEACON) to assess self-reported risk for common mental disorders among children and adolescents from grades 4 through 11 (approximate ages 9 through 17). Replicated results from two large cohorts demonstrate strong overall psychometric properties. Items were developed to assess four common forms of mental disorder—anxiety, depression, attention/hyperactivity problems, and eating difficulties. IRT results demonstrated that all items provided good discrimination across their relevant construct. These results were supported by the strong item-total correlations showed by each item with its relevant subscale, along with the subsequent indicators of subscale reliability. Construct validity was demonstrated by moderate to strong correlations between each BEACON subscale and its parallel, self-reported score on widely

**TABLE 7** Study 2 results for analyses derived from item response theory and classical test theory for final items for grades 6–11.

	Results from item response theory		Results from classical test theory					Item-total correlation	
	Discrimination parameter	DIFF (gender)	Good ICCs slope and spread	4 CCC crossovers	Mean	SD	Skew		Kurtosis
<b>Anxiety (10 items)</b>									
Worried about things	2.97	AA	Yes	Yes	2.88	1.16	0.05	-0.83	0.76
Felt afraid	2.81	AA	Yes	Yes	1.96	1.05	0.93	0.16	0.75
Afraid of making a fool of myself	2.19	AA	Yes	Yes	2.48	1.32	0.48	-0.94	0.70
Scared for no reason at all	2.90	AA	Yes	Yes	1.87	1.16	1.19	0.36	0.74
Felt nervous	2.73	AA	Yes	Yes	2.61	1.23	0.33	-0.88	0.75
Bothered you when you felt scared or worried	2.96	AA	Yes	Yes	2.41	1.14	0.44	-0.74	0.77
Hard to get on with parents and family	1.98	AA	Yes	Yes	1.98	1.17	1.01	0.01	0.68
Hard to get on with friends	1.80	AA	Yes	Yes	1.80	1.04	1.25	0.79	0.63
Hard to look after your health	2.36	AA	Yes	Yes	2.17	1.30	0.80	-0.59	0.73
Hard to do activities you enjoy	2.13	AA	Yes	Yes	1.91	1.17	1.14	0.23	0.69
<b>Depression (9 items)</b>									
Felt really sad	3.60	AA	Yes	Yes	18.17	9.17	0.96	-0.02	$\alpha = 0.95$
Felt really alone	3.22	AA	Yes	Yes	2.09	1.17	0.86	-0.22	0.83
Felt like there is nothing to look forward to	2.43	AA	Yes	Yes	2.01	1.24	1.00	-0.16	0.80
Didn't enjoy things anymore	2.79	AA	Yes	Yes	2.06	1.28	0.95	-0.32	0.75
Felt like crying a lot	3.02	BB-	Yes	Yes	1.92	1.19	1.13	0.20	0.78
Bothered you when you felt sad or lonely	3.32	AA	Yes	Yes	2.00	1.33	1.09	-0.14	0.77
					2.26	1.21	0.65	-0.64	0.80

(Continues)

TABLE 7 (Continued)

	Results from item response theory			Results from classical test theory					Item-total correlation
	Discrimination parameter	DIFF (gender)	Good ICCs slope and spread	4 CCC crossovers	Mean	SD	Skew	Kurtosis	
Hard to look after your health	3.38	AA	Yes	Yes	1.99	1.21	1.03	-0.01	0.81
Hard to do daily activities	2.96	AA	Yes	Yes	2.04	1.23	0.95	-0.20	0.78
Hard to do activities you enjoy	3.09	AA	Yes	Yes	1.81	1.12	1.28	0.70	0.78
Attention/hyperactivity problems (8 items)									
Got distracted easily	3.07	AA	Yes	Yes	18.09	7.94	0.63	-0.47	$\alpha = 0.92$
Hard to manage my time properly	2.62	AA	Yes	Yes	2.76	1.31	0.22	-1.11	0.78
Didn't enjoy tasks where had to concentrate	2.75	AA	Yes	Yes	2.61	1.31	0.35	-1.03	0.74
Hard to be still when needed to be	2.48	AA	Yes	Yes	2.57	1.36	0.37	-1.11	0.76
Bother you when you couldn't focus or stay still	2.58	AA	Yes	Yes	2.26	1.36	0.74	-0.75	0.72
Hard to look after your health	2.58	AA	Yes	Yes	2.38	1.23	0.52	-0.81	0.74
Hard to do daily activities	2.90	AA	Yes	Yes	1.78	1.08	1.31	0.84	0.71
Hard to do activities you enjoy	2.26	AA	Yes	Yes	2.10	1.22	0.88	-0.30	0.77
Eating difficulties (8 items)									
Watched what I ate so I wouldn't gain weight	2.16	AA	Yes	Yes	14.69	8.00	1.28	0.72	$\alpha = 0.93$
Terrified of gaining weight	3.42	AA	Yes	Yes	2.49	1.06	0.33	-0.52	0.72
Thought I'm fatter than other people think I am	3.28	AA	Yes	Yes	1.96	1.00	0.92	0.38	0.69
After eating, felt guilty and upset with myself	4.04	AA	Yes	Yes	2.19	1.25	0.77	-0.51	0.60
					1.76	1.11	1.37	0.86	0.66

TABLE 7 (Continued)

	Results from item response theory		Results from classical test theory						
	Discrimination parameter	DIFF (gender)	Good ICCs slope and spread	4 CCC crossovers	Mean	SD	Skew	Kurtosis	Item-total correlation
Hard to get on with parents and family members	3.33	AA	Yes	Yes	1.58	1.05	1.81	2.33	0.75
Hard to look after your health	4.03	AA	Yes	Yes	1.76	1.17	1.47	1.03	0.81
Hard to do daily activities	3.69	AA	Yes	Yes	1.52	0.97	1.97	3.15	0.76
Hard to do activities you enjoy	3.62	AA	Yes	Yes	1.56	1.03	1.88	2.67	0.76

Abbreviations: CCC, category characteristic curve; ICC, item characteristic curve.

TABLE 8 Correlations between BEACON subscales and validator measures for grades 6–11.

	BEACON—total	BEACON—anxiety	BEACON—depression	BEACON—attention/hyperactivity problems	BEACON—eating difficulties
Student self-report					
BEACON—anxiety (n = 9404)	0.92**				
BEACON—depression (n = 9312)	0.94**	0.86**			
BEACON—attention/hyperactivity problems (n = 9270)	0.86**	0.75**	0.76**		
BEACON—Eating Difficulties (n = 9239)	0.81**	0.64**	0.67**	0.58**	
Peer victimization (n = 7242)	0.46**	0.41**	0.42**	0.39**	0.39**
RADAR Youth Version—School Connectedness Subscale (n = 7050)	-0.47**	-0.43**	-0.46**	-0.41**	-0.35**
RADAR Youth Version—Family Relations Subscale (n = 7046)	-0.41**	-0.36**	-0.41**	-0.35**	-0.32**
RADAR Youth Version—Peer Acceptance Subscale (n = 7051)	-0.43**	-0.42**	-0.42**	-0.37**	-0.32**
Total number of services seen (n = 9405)	0.27**	0.26**	0.26**	0.23**	0.20**
Parent report and school data					
BEACON Parent Version—anxiety (n = 88)	0.42**	0.43**	0.36**	0.42**	0.25*
BEACON Parent Version—depression (n = 87)	0.24*	0.25*	0.24*	0.28**	0.05
BEACON Parent Version—attention/hyperactivity problems (n = 87)	0.25*	0.25*	0.21*	0.33**	0.05
BEACON Parent Version—eating difficulties (n = 87)	0.26*	0.19	0.18	0.15	0.35**
Total number of services seen by their child (n = 87)	0.30**	0.27*	0.25*	0.31**	0.20
Mean school grades (n = 137)	-0.19*	-0.12	-0.22*	-0.17*	-0.16
Total absences from school (n = 137)	0.18*	0.14	0.16*	0.22**	0.14

\*\* $p < .01$ , \* $p < .05$ .

**TABLE 9** Differences between students at high risk for suicidality and self-harm and rest of sample on validation measures (grades 7 and above only).

	High suicidality risk (n = 187)	Rest of sample (n = 7426)	Statistics	High self-harm risk (n = 149)	Rest of sample (n = 7464)	Statistics
BEACON—anxiety (n = 7586)	34.96 (9.56)	21.69 (8.94)	F(1, 7584) = 400.77, p < .001, d = 1.48	34.39 (9.19)	21.77 (9.01)	F(1, 7584) = 286.26, p < .001, d = 1.40
BEACON—depression (n = 7501)	34.16 (8.69)	17.75 (8.84)	F(1, 7499) = 629.18, p < .001, d = 1.86	33.77 (8.46)	17.84 (8.93)	F(1, 7499) = 464.80, p < .001, d = 1.78
BEACON—attention/hyperactivity problems (n = 7462)	29.01 (7.75)	17.97 (7.81)	F(1, 7460) = 364.95, p < .001, d = 1.42	28.50 (7.46)	18.04 (7.87)	F(1, 7460) = 259.11, p < .001, d = 1.33
BEACON—eating difficulties (n = 7437)	26.77 (9.79)	14.27 (7.73)	F(1, 7435) = 469.66, p < .001, d = 1.61	25.71 (9.77)	14.36 (7.83)	F(1, 7435) = 303.52, p < .001, d = 1.44
Total number of services seen (n = 7588)	0.82 (1.33)	0.22 (0.67)	F(1, 7586) = 136.99, p < .001, d = .87	0.79 (1.32)	0.23 (0.67)	F(1, 7586) = 94.31, p < .001, d = .80
Mean school grades (n = 104)	2.50 (0.71)	3.67 (0.68)	F(1, 102) = 5.83, p = .018, d = 1.72	2.74 (0.65)	3.67 (0.68)	F(1, 102) = 5.50, p = .021, d = 1.37

used, standard measures (e.g., depression, ADHD). Significant correlations were also shown with some parent-reported scores on standard measures as well as a parent-report version of the BEACON, although as expected due to the different reporters (De Los Reyes & Kazdin, 2005), these correlations were considerably smaller.

The BEACON was designed to assess several common forms of mental disorder found among children and adolescents. No screener can assess for all forms of disorder without becoming prohibitively lengthy and therefore key decisions had to be made about its coverage. In our initial discussions, we decided to not assess oppositional and conduct problems since these types of problems are characterized by overt impact on others and are well detected by parents and teachers (Heerde & Hemphill, 2018; Phares & Danforth, 1994) and children and adolescents report little personal distress about these behaviors (Phares & Danforth, 1994). Similarly, the illicit nature of substance use at these ages meant that screening for alcohol and other substance abuse raised concerns about honesty as well as additional concerns of legal reporting requirements. Ultimately, the final version of the BEACON assesses the most common forms of mental disorder among children and adolescents (aside from oppositional/conduct problems) (Lawrence et al., 2015; Polanczyk et al., 2015). Consistent with age of onset data patterns and age-related prevalence data (Kessler et al., 2007; Lawrence et al., 2015; Rapee et al., 2019), two versions of the BEACON were developed. The version for younger cohorts (grades 4 and 5, ages around 9–11 years) assessed only attention/hyperactivity problems and anxiety, while the version for older cohorts (grades 6 and above; ages 11+ years) added subscales assessing eating difficulties and depression. Future additions could add items assessing additional constructs, although this would come at the cost of additional length.

At the request of schools, who wanted to be able to identify any young people who were at immediate risk, we included items to assess self-harm and suicidality. These items were included on a voluntary basis by schools, depending on their judgment of acceptance by the community. However, within our study, the majority of eligible schools opted to include them and there was no received complaint from a parent about these items. The small identified at-risk sample (2.5% at-risk for suicidality and 2.0% at-risk for self-harm), along with the nondimensional nature of the scales, meant that formal psychometric properties were not evaluated. However, the items had good clinical, face-validity and the identified adolescents showed marked distress and impairment on relevant measures (other BEACON subscales and school grades).

A unique feature of the BEACON was its inclusion of items to assess functional impairment related to each distinct mental health difficulty. Most definitions of mental disorder include an assumption of associated life impairment (American Psychiatric Association, 2013; Polanczyk et al., 2015; Rapee et al., 2012). Hence, including items assessing impairment should provide a better representation of a clinical disorder than assessment of symptoms alone, as is found in most current questionnaire measures. Contrary to this suggestion, in our preliminary analyses, we found very strong correlations between the symptom and impairment items from each subscale. Further, as reported in the results, item-total correlations for both the symptom and impairment items were strong and similar in magnitude, indicating that they both assessed a consistent construct and that the impairment items most likely did not provide extensive unique information. This result is somewhat surprising given evidence from epidemiological studies that adding impairment criteria to diagnostic decisions markedly impacts estimated prevalence over data derived from symptoms alone (Bird et al., 1990; Costello, 1996). Future research is needed to determine whether a measure comprising both symptom and impairment items provides better identification of children and adolescents with mental disorder than a measure comprised of only symptoms.

An interesting conceptual issue is the relationship between BEACON scores and prior help-seeking among children and adolescents (and their carers). Our data from study 2 indicated significant, but small relationships between BEACON subscales and number of services seen by the child. The significant relationship can be seen as further evidence of validity—those who score higher on the BEACON are more likely to have received help for mental health. However, the small effect sizes indicate a less than perfect relationship—that is, many children and adolescents scoring high on the BEACON have not previously received help. Partly this effect may be a result of errors in measurement—both on the BEACON and especially our broad measure of help-seeking (number of services). But the effect is not unsurprising given extensive evidence that many children and adolescents with

mental disorders do not receive adequate help (Ford et al., 2007; Merikangas et al., 2011; Reardon et al., 2020). In many ways, the low correlations indicate the potential value of the BEACON in identifying children and adolescents who may potentially slip between the cracks.

We refrain from describing specific cut-off scores since relevant cut scores may differ depending on the aims of the screening and in the current study we do not have a “gold standard.” There are also no existing “objective markers” of mental disorder as there might be, for example, in fields such as oncology. Even structured diagnostic interviews have very limited reliability and, therefore, limited validity. We therefore suggest that the *T*-scores described in the Supporting Information tables provide users of the BEACON with guidelines to relevant cut points to identify students who might benefit from further evaluation or referral. Naturally, the specific cut-score selected by a school or professional will vary based on objectives of the screening and the school's specific circumstances. For example, to identify the approximate proportion of the population expected to be diagnosed with each relevant mental disorder (Lawrence et al., 2015; Nagl et al., 2016; Polanczyk et al., 2015), a school might use cut-offs at *T*-scores of 67 for grades 4–5 and 66 for grades 6–11 for anxiety and attention/hyperactivity problems (reflecting the highest 7%) and for the older grades, 73 for depression and 74 for eating difficulties (reflecting the highest 3%). However, if a school has very limited local resources, then they might select an even higher cut-off to reduce the identified numbers (accepting that this will increase false negatives), or if they plan to implement broad, early intervention programs or if they wish to be more inclusive and err toward more false positives but fewer false negatives, then a lower cut-off may be warranted.

Several limitations of the research require some discussion. The key limitation is the lack of a gold standard measure of actual mental disorder against which to determine sensitivity and specificity. However, obtaining a perfect gold standard is probably not possible. Structured diagnostic interviews are the most widely used standard, but they themselves are subject to limited reliability and hence limitations on their validity (De Los Reyes et al., 2015; McLellan et al., 2021; Pelham et al., 2005). Nonetheless, future research would benefit from comparing scores on the BEACON against some form of clinical indicator, although as noted above, measures such as help-seeking are not ideal as an indicator of validity. It should also be noted that several of our current validation measures, such as absenteeism and parent-reported symptoms were derived through opt-in consent and may, therefore, represent biased samples. As noted previously, the lack of coverage of oppositional or conduct problems limits the thoroughness of the BEACON. However, these disorders are less likely to be amenable to self-report and may require parent or teacher report for their identification. It should also be noted that this research occurred during the height of the COVID period in Australia and many communities locked down at various stages. Hence the data, and especially the norms, may not represent the typical functioning of children and adolescents. The large sample of relatively representative children and adolescents, including representation of broad socioeconomic backgrounds, is a strength of this study. However, by definition, this representation is dominated by the dominant culture. Hence although represented within this sample, the proportions of low-frequency populations (such as non-English-speaking background, Indigenous, and disadvantaged background) were not large enough to test psychometric properties separately. Given differences in proportions of and attitudes toward mental disorder among these populations, it is possible that the psychometric properties of the BEACON might differ. Hence, future research that over-samples from these sub-populations is warranted to assess psychometric properties specifically within these groups.

In summary, we have developed a self-report instrument that allows identification of children and adolescents from around 9 to 17 years who are at significant risk of diagnosable mental disorders. It needs to be remembered that identifying mental disorder among school students carries significant moral and ethical considerations and anyone who plans to use this measure needs to ensure careful deliberation of these issues. In our work, we developed detailed guidelines for schools about practical and ethical ways to deliver the screening process and to guide school personnel about ways to provide information to families (Rapee et al., 2024). Similar guidelines were developed for families about ways to interpret high scores and potential sources of local help. Pre-screening consultation with local mental health services is critical to ensure smooth pathways to care (Burns & Rapee, 2021).



The BEACON is not a diagnostic or stand-alone mental health tool and some research has pointed to the value of an additional, second screening step or the inclusion of multiple informants (Goodman et al., 2003; Scott et al., 2009). In brief, the BEACON provides a first step in helping to facilitate early guidance and help for children and adolescents who may be struggling with mental health difficulties and allowing them a path to positive future development.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the generous support of the BUPA Health Foundation for funding this research. Open access publishing facilitated by Macquarie University, as part of the Wiley - Macquarie University agreement via the Council of Australian University Librarians.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings will be available following an embargo from the date of publication to allow for utilization of all research findings.

## ORCID

Ronald M. Rapee  <http://orcid.org/0000-0002-1724-1076>

## PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/jclp.23673>.

## ENDNOTE

<sup>1</sup> Unfortunately, given the very small numbers, participants who identified as a gender other than male or female had to be excluded from this analysis.

## REFERENCES

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.
- Anderson, J. K., Ford, T., Soneson, E., Coon, J. T., Humphrey, A., Rogers, M., Moore, D., Jones, P. B., Clarke, E., & Howarth, E. (2019). A systematic review of effectiveness and cost-effectiveness of school-based identification of children and young people at risk of, or currently experiencing mental health difficulties. *Psychological Medicine*, 49(1), 9–19. <https://doi.org/10.1017/S0033291718002490>
- Angold, A., Costello, E. J., Farmer, E. M., Burns, B. J., & Erkanli, A. (1999). Impaired but undiagnosed. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38(2), 129–137. <https://doi.org/10.1097/00004583-199902000-00011>
- Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., & Silver, D. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, 5, 237–249.
- Australian Bureau of Statistics. (2016a). *Australian Statistical Geography Standard (ASGS): Volume 5—Remoteness Structure*. Australian Bureau of Statistics. (2016b). *Census of population and housing: Socio-economic indexes for areas (SEIFA)*. <https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001>
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.
- Bird, H. R., Yager, T. J., Staghezza, B., Gould, M. S., Canino, G., & Rubio-Stipec, M. (1990). Impairment in the epidemiological measurement of childhood psychopathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29(5), 796–803.
- Bruhn, A. L., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional and behavioral screening practices in K-12 schools. *Education & Treatment of Children*, 37(4), 611–634. <https://doi.org/10.1353/etc.2014.0039>

- Burns, J. R., & Rapee, R. M. (2016). Screening for mental health risk in high schools: The development of the Youth RADAR. *Psychological Assessment, 28*(10), 1220–1231. <https://doi.org/10.1037/pas0000237>
- Burns, J. R., & Rapee, R. M. (2019). School-based assessment of mental health risk in children: The preliminary development of the child RADAR. *Child and Adolescent Mental Health, 24*(1), 66–75. <https://doi.org/10.1111/camh.12258>
- Burns, J. R., & Rapee, R. M. (2021). From barriers to implementation: Advancing universal mental health screening in schools. *Journal of Psychologists and Counsellors in Schools, 31*(2), 172–183. <https://doi.org/10.1017/jgc.2021.17>
- Burns, J. R., & Rapee, R. M. (2022). Barriers to universal mental health screening in schools: The perspective of school psychologists. *Journal of Applied School Psychology, 38*(3), 223–240. <https://doi.org/10.1080/15377903.2021.1941470>
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Lawrence Erlbaum Associates.
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy, 38*, 835–855.
- Costello, E. J. (1996). The Great Smoky Mountains Study of Youth: Functional impairment and serious emotional disturbance. *Archives of General Psychiatry, 53*(12), 1137–1143. <https://doi.org/10.1001/archpsyc.1996.01830120077013>
- Creswell, C., Nauta, M. H., Hudson, J. L., March, S., Reardon, T., Arendt, K., Bodden, D., Cobham, V. E., Donovan, C., Halldorsson, B., In-Albon, T., Ishikawa, S.-I., Johnsen, D. B., Jolstedt, M., de Jong, R., Kreuze, L., Mobach, L., Rapee, R. M., Spence, S. H., ... Kendall, P. C. (2021). Research Review: Recommendations for reporting on treatment trials for child and adolescent anxiety disorders—an international consensus statement. *Journal of Child Psychology and Psychiatry, 62*(3), 255–269. <https://doi.org/10.1111/jcpp.13283>
- Deighton, J., Tymms, P., Vostanis, P., Belsky, J., Fonagy, P., Brown, A., Martin, A., Patalay, P., & Wolpert, M. (2013). The development of a school-based measure of child mental health. *Journal of Psychoeducational Assessment, 31*(3), 247–257. <https://doi.org/10.1177/0734282912465570>
- Dickson, S. J., Kuhnert, R.-L., Lavell, C. H., & Rapee, R. M. (2022). Impact of psychotherapy for children and adolescents with anxiety disorders on global and domain-specific functioning: A systematic review and meta-analysis. *Clinical Child and Family Psychology Review, 25*(4), 720–736. <https://doi.org/10.1007/s10567-022-00402-7>
- Dowdy, E., Dever, B. V., Raines, T. C., & Moffa, K. (2016). A preliminary investigation into the added value of multiple gates and informants in universal screening for behavioral and emotional risk. *Journal of Applied School Psychology, 32*(2), 178–198. <https://doi.org/10.1080/15377903.2016.1165327>
- DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *DHD Rating Scale-IV: Checklists, norms, and clinical interpretation*. Guilford. <https://doi.org/10.1177/0734282905285792>
- Evans, R., Thirlwall, K., Cooper, P., & Creswell, C. (2017). Using symptom and interference questionnaires to identify recovery among children with anxiety disorders. *Psychological Assessment, 29*(7), 835–843. <https://doi.org/10.1037/pas0000375>
- Fairburn, C. G., & Cooper, Z. (1993). The eating disorder examination. In C. G. Fairburn, & G. T. Wilson (Eds.), *Binge Eating: Nature, Assessment and Treatment* (12th ed., pp. 317–360). Guilford Press.
- Ford, T., Hamilton, H., Meltzer, H., & Goodman, R. (2007). Child mental health is everybody's business: The prevalence of contact with public sector services by type of disorder among British school children in a three-year period. *Child and Adolescent Mental Health, 12*, 13–20. <https://doi.org/10.1111/j.1475-3588.2006.00414.x>
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2003). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry, 15*(1–2), 166–172. <https://doi.org/10.1080/0954026021000046128>
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Heerde, J. A., & Hemphill, S. A. (2018). Examination of associations between informal help-seeking behavior, social support, and adolescent psychosocial outcomes: A meta-analysis. *Developmental Review, 47*, 44–62. <https://doi.org/10.1016/j.dr.2017.10.001>
- Herman, K. C., Reinke, W. M., Huang, F. L., Thompson, A. M., & Doyle-Barker, L. (2021). An investigation of the psychometric properties of the early identification system-student report in a middle school sample. *School Psychology, 36*(1), 34–46. <https://doi.org/10.1037/spq0000414>
- Hunt, C., Peters, L., & Rapee, R. M. (2012). Development of a measure of the experience of being bullied in youth. *Psychological Assessment, 24*(1), 156–165. <https://doi.org/10.1037/a0025178>
- IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0*. IBM Corp.
- Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., Setmeyer, A. J., Rodelo, C., Newcomb-McNeal, E., & Tennant, J. (2014). A critical review of five commonly used social-emotional and behavioral screeners

- for elementary or secondary schools. *Contemporary School Psychology*, 18(4), 241–254. <https://doi.org/10.1007/s40688-014-0026-6>
- Kamphaus, R. W., Reynolds, C. R., & Dever, B. V. (2014). Behavioral and mental health screening. In R. J. Kettler, T. A. Glover, C. A. Albens & K. A. Feeney-Kettler, (Eds.), *Universal Screening in Educational Settings: Evidence-Based Decision Making for Schools* (pp. 249–273). American Psychological Association. <https://doi.org/10.1037/14316-010>
- Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & ?st??n, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, 20, 359–364.
- Kilgus, S. P., von der Embse, N. P., Taylor, C. N., Van Wie, M. P., & Sims, W. A. (2018). Diagnostic accuracy of a universal screening multiple gating procedure: A replication study. *School Psychology Quarterly*, 33(4), 582–589. <https://doi.org/10.1037/spq0000246>
- Lawrence, D., Johnson, S., Hafekost, J., Boterhoven De Haan, K., Sawyer, M. G., Ainley, J., & Zubrick, S. (2015). *The Mental Health of Children and Adolescents. Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing*. Department of Health.
- Levitt, J. M., Saka, N., Hunter Romanelli, L., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45, 163–191. <https://doi.org/10.1016/j.jsp.2006.11.005>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858–900. <https://doi.org/10.1037/a0038498>
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131(4), 483–509. <https://doi.org/10.1037/0033-2909.131.4.483>
- De Los Reyes, A., Makol, B. A., Racz, S. J., Youngstrom, E. A., Lerner, M. D., & Keeley, L. M. (2019). The work and social adjustment scale for youth: A measure for assessing youth psychosocial impairment regardless of mental health status. *Journal of child and family studies*, 28(1), 1–16. <https://doi.org/10.1007/s10826-018-1238-6>
- Lyneham, H. J., Sbrulati, E. S., Abbott, M. J., Rapee, R. M., Hudson, J. L., Tolin, D. F., & Carlson, S. E. (2013). Psychometric properties of the Child Anxiety Life Interference Scale (CALIS). *Journal of Anxiety Disorders*, 27(7), 711–719. <https://doi.org/10.1016/j.janxdis.2013.09.008>
- McLellan, L. F., Kangas, M., Rapee, R. M., Iverach, L., Wuthrich, V. M., Hudson, J. L., & Lyneham, H. J. (2021). The Youth Online Diagnostic Assessment (YODA): Validity of a new tool to assess anxiety disorders in youth. *Child Psychiatry and Human Development*, 52, 270–280. <https://doi.org/10.1007/s10578-020-01007-3>
- Merikangas, K. R., He, J.-p, Burstein, M., Swendsen, J., Avenevoli, S., Case, B., Georgiades, K., Heaton, L., Swanson, S., & Olfson, M. (2011). Service utilization for lifetime mental disorders in U.S. adolescents: Results of the National comorbidity Survey-Adolescent Supplement (NCSA). *Journal of the American Academy of Child and Adolescent Psychiatry*, 50(1), 32–45.
- Meyer, J. P. (2014). *Applied measurement with jMetrik* (1st ed.). Routledge. <https://doi.org/10.4324/9780203115190>
- Moore, S. A., Widales-Benitez, O., Carnazzo, K. W., Kim, E. K., Moffa, K., & Dowdy, E. (2015). Conducting universal complete mental health screening via student self-report. *Contemporary School Psychology*, 19, 253–267. <https://doi.org/10.1007/s40688-015-0062-x>
- Nagl, M., Jacobi, C., Paul, M., Beesdo-Baum, K., Höfler, M., Lieb, R., & Wittchen, H. U. (2016). Prevalence, incidence, and natural course of anorexia and bulimia nervosa among adolescents and young adults. *European Child & Adolescent Psychiatry*, 25, 903–918.
- Nauta, M. H., Scholing, A., Rapee, R. M., Abbott, M., Spence, S. H., & Waters, A. (2004). A parent-report measure of children's anxiety: Psychometric properties and comparison with child-report in a clinic and normal sample. *Behaviour Research and Therapy*, 42, 813–839. [https://doi.org/10.1016/S0005-7967\(03\)00200-6](https://doi.org/10.1016/S0005-7967(03)00200-6)
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Pelham, Jr., W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 34(3), 449–476. [https://doi.org/10.1207/s15374424jccp3403\\_5](https://doi.org/10.1207/s15374424jccp3403_5)
- Phares, V., & Danforth, J. S. (1994). Adolescents', parents', and teachers' distress over adolescents' behavior. *Journal of Abnormal Child Psychology*, 22, 721–732. <https://doi.org/10.1007/BF02171998>
- Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, 56(3), 345–365. <https://doi.org/10.1111/jcpp.12381>
- Productivity Commission. (2020). *Mental health* (Report no. 95). Australian Government.

- Rapee, R. M., Bögels, S. M., van der Sluis, C. M., Craske, M. G., & Ollendick, T. (2012). Annual research review: Conceptualising functional impairment in children and adolescents. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 53, 454–468. <https://doi.org/10.1111/j.1469-7610.2011.02479.x>
- Rapee, R. M., Kuhnert, R.-L., Bowsher, I., Burns, J., Coen, J., Dixon, J., Kotselas, P., Lourey, C., McLellan, L. F., Mihalopoulos, C., Peters, L., Roos, T., Thomas, D., & Wuthrich, V. (2024). A randomised controlled trial of the impact of school-based, universal, mental health screening and feedback on students' mental health and help-seeking. Manuscript submitted for publication.
- Rapee, R. M., Oar, E. L., Johnco, C. J., Forbes, M. K., Fardouly, J., Magson, N. R., & Richardson, C. E. (2019). Adolescent development and risk for the onset of social-emotional disorders: A review and conceptual model. *Behaviour Research and Therapy*, 123, 103501. <https://doi.org/10.1016/j.brat.2019.103501>
- Reardon, T., Harvey, K., & Creswell, C. (2020). Seeking and accessing professional support for child anxiety in a community sample. *European Child & Adolescent Psychiatry*, 29, 649–664. <https://doi.org/10.1007/s00787-019-01388-4>
- Renshaw, T. L., & Cook, C. R. (2018). Initial development and validation of the youth internalizing problems screener. *Journal of Psychoeducational Assessment*, 36(4), 366–378. <https://doi.org/10.1177/0734282916679757>
- Romer, D., & McIntosh, M. (2005). The roles and perspectives of school mental health professionals in promoting adolescent mental health. In D. L. Evans, E. B. Foa, R. E. Gur, H. Hendin, C. P. O'Brien, M. E. P. Seligman, & B. T. Walsh (Eds.), *Treating and preventing adolescent mental health disorders: What we know and what we don't know. A research agenda for improving the mental health of our youth*. Oxford University Press. <https://doi.org/10.1093/9780195173642.003.0032>
- Scott, M. A., Wilcox, H. C., Schonfeld, I. S., Davies, M., Hicks, R. C., Turner, J. B., & Shaffer, D. (2009). School-based screening to identify at-risk students not already known to school professionals: The Columbia Suicide Screen. *American Journal of Public Health*, 99(2), 334–339. <https://doi.org/10.2105/AJPH.2007.127928>
- Soneson, E., Childs-Fegredo, J., Anderson, J. K., Stochl, J., Fazel, M., Ford, T., Humphrey, A., Jones, P. B., & Howarth, E. (2018). Acceptability of screening for mental health difficulties in primary schools: A survey of UK parents. *BMC Public Health*, 18(1), 1404. <https://doi.org/10.1186/s12889-018-6279-7>
- Soneson, E., Howarth, E., Ford, T., Humphrey, A., Jones, P. B., Thompson Coon, J., Rogers, M., & Anderson, J. K. (2020). Feasibility of school-based identification of children and adolescents experiencing, or at-risk of developing, mental health difficulties: A systematic review. *Prevention Science*, 21, 581–603. <https://doi.org/10.1007/s11121-020-01095-6>
- Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, 36(5), 545–566. [https://doi.org/10.1016/S0005-7967\(98\)00034-5](https://doi.org/10.1016/S0005-7967(98)00034-5)
- Spence, S. H., & Rapee, R. M. (2022). The development and preliminary validation of a brief scale of emotional distress in young people using combined classical test theory and item response theory approaches: The Brief Emotional Distress Scale for Youth (BEDSY). *Journal of Anxiety Disorders*, 85, 102495. <https://doi.org/10.1016/j.janxdis.2021.102495>
- StataCorp. (2021). *Stata statistical software: Release 17*. StataCorp LLC.
- Stevens, K. J. (2010). Working with children to develop dimensions for a preference-based, generic, pediatric health-related quality-of-life measure. *Qualitative Health Research*, 20, 340–351.
- Williamson, V., Larkin, M., Reardon, T., Pearcey, S., Button, R., Green, I., Hill, C., Stallard, P., Spence, S. H., Breen, M., McDonald, I., Ukoumunne, O., Ford, T., Violato, M., Sniehotta, F., Stainer, J., Gray, A., Brown, P., Sancho, M., ... Creswell, C. (2022). School-based screening for childhood anxiety problems and intervention delivery: A codesign approach. *BMJ Open*, 12(6), e058089. <https://doi.org/10.1136/bmjopen-2021-058089>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Rapee, R. M., Kuhnert, R., Spence, S. H., Bowsher, I., Burns, J., Coen, J., Dixon, J., Kotselas, P., Lourey, C., McLellan, L. F., Mihalopoulos, C., Peters, L., Prendergast, T., Roos, T., Thomas, D., & Wuthrich, V. (2024). The Brief Evaluation of Adolescents and Children Online (BEACON): Psychometric development of a mental health screening measure for school students. *Journal of Clinical Psychology*, 80, 1420–1447. <https://doi.org/10.1002/jclp.23673>