ORIGINAL CONTRIBUTION

Mental health matters in elementary school: first-grade screening predicts fourth grade achievement test scores

Maria Paz Guzman · Michael Jellinek · Myriam George · Marcela Hartley ·
Ana Maria Squicciarini · Katia M. Canenguez · Karen A. Kuhlthau ·
Recai Yucel · Gwyne W. White · Javier Guzman · J. Michael Murphy

Abstract The objective of the study was to evaluate whether mental health problems identified through screens administered in first grade are related to poorer academic achievement test scores in the fourth grade. The government of Chile uses brief teacher- and parent-completed measures [Teacher Observation of Classroom Adaptation-Revised (TOCA-RR) and Pediatric Symptom Checklist (PSC-Cl)] to screen for mental health problems in about one-fifth of the country's elementary schools. In fourth grade, students take the national achievement tests (SIMCE) of language, mathematics and science. This study examined whether mental health problems identified through either or both screens predicted achievement test scores after controlling for student and family risk factors. A total of 17,252 students had complete first grade teacher forms and these were matched with fourth grade SIMCE data for 11,185 students, 7,903 of whom also had complete parent form data from the first grade. Students at risk on either the TOCA-RR or the PSC-Cl or both performed significantly worse on all SIMCE subtests. Even after controlling for covariates and adjusting for missing data, students with mental health problems on one screen in first grade had

M. P. Guzman · M. George · M. Hartley ·
A. M. Squicciarini · J. Guzman
Department of School Health,
Junta Nacional de Auxillo Escolar y Becas, Santiago, Chile
e-mail: mariapaz@americasolidaria.org

M. George
e-mail: mpgeorge@junaeb.cl

A. M. Squicciarini
e-mail: asquicciarini@junaeb.cl

J. Guzman
e-mail: jguzman@junaeb.cl

M. George
Faculty of Psychology, University of Chile, Santiago, Chile

M. Jellinek · J. M. Murphy
Department of Psychiatry, Massachusetts General Hospital,
Boston, USA
e-mail: mjellinek@partners.org

M. Jellinek · J. M. Murphy
Department of Psychiatry, Harvard Medical School,
Boston, USA

M. Jellinek
Newton Wellesley Hospital, Newton, USA

K. M. Canenguez
Department of Psychology, University of Massachusetts,
Boston, USA
e-mail: kcanenguez@partners.org

K. A. Kuhlthau
Department of Pediatrics, Massachusetts General Hospital,
Boston, USA
e-mail: kkuhlthau@partners.org

R. Yucel
Department of Epidemiology and Biostatistics,
State University of New York at Albany, New York, USA
e-mail: ryucel@uamail.albany.edu

G. W. White
Department of Psychology,
Rutgers, The State University of New Jersey,
New Jersey, USA
e-mail: gwyneage@gmail.com

J. M. Murphy (✉)
Child Psychiatry Service, Massachusetts General Hospital,
Yawkey 6a, Boston, MA 02114, USA
e-mail: mmurphy6@partners.org

fourth grade achievement scores that were 14–18 points (~1/3 SD) lower than students screened as not at risk. Students at risk on both screens had scores that were on average 33 points lower than students at risk on either screen. Mental health problems in first grade were one of the strongest predictors of lower achievement test scores 3 years later, supporting the premise that for children mental health matters in the real world.

Keywords Child mental health · Psychosocial screening · Academic achievement · Risk factors

Introduction

Although for several decades there has been widespread, international agreement on the importance of expanding the focus of pediatric primary care to include mental as well as physical health problems and of implementing routine screening for psychosocial problems as a strategy for achieving more integrated care [23, 48], it is only in the past few years that several states and countries have established large-scale screening programs [29, 41]. In the USA, for example, although for the past two decades the nation's health goals [45, 46] have recommended routine screening for mental health problems as a part of primary care for both children and adults and although for the past decade court orders in a number of states have reaffirmed the requirement for screening of children from low-income families who have government health insurance [29, 42], it was not until 2008 that the first state in the USA made screening a formal requirement [18, 40].

The prevalence and impact of children's mental health problems outside the USA have also long been recognized in both developed and developing countries. Over the past three decades, the World Health Organization has consistently and repeatedly advocated for the inclusion of enhanced mental health screening and treatment programs in primary health care and in schools [12, 19, 52]. Studies of the global burden of mental health problems have been widely accepted and are influential [35, 48] and applied to children as well as adults [52]. Over this period, there have been many studies of children's mental health problems and interventions around the world. The Pediatric Symptom Checklist has been translated into two dozen languages and validated in Japan [20], Holland [41], Austria [44], the Philippines [7] and Brazil [36], as well as Chile [10]. The Strengths and Difficulties Questionnaire, another brief screen like the PSC, has been translated into 50 languages and the Child Behavior Checklist into even more [50]. Despite the widespread concern about children's mental health throughout the world over many decades, only a single country, the Netherlands [41], mandates mental

health screening for all children. In Chile, screening is done in only some designated high-risk schools [17] and in the USA only a single state enforces its screening requirement and only for children with public health insurance [29].

Although many would agree that screening for earlier recognition of psychosocial dysfunction has face validity, data on the implications of a positive mental health screen for children's overall adjustment, health and academic achievement are limited. This gap is evident in the US Preventive Services Task Force's recent decision to endorse routine screening for depression (in adolescents), but not routine screening for overall mental health because of the lack of evidence linking screened risk to real-world outcomes and improved functioning after intervention [47]. Nevertheless, the needs are great enough and the face validity of interventions strong enough that recent position papers by the American Academy of Pediatrics recommend routine screening especially in settings where services are readily available [4, 5].

Screening for depression is an example of what is called 'narrow band' screening, while screening for overall psychosocial functioning or mental health is called 'broadband' screening [37]. There have been many studies that document the relationship between academic outcomes and specific diagnoses like depression [13], diagnosis specific measures like the Children's Depression Inventory [2], domain-specific problems like accepting a teacher's authority [39] and general conditions like serious emotional disturbance [49]. But, only one of the narrow band studies mentioned above [49] used the kind of standardized academic achievement tests of reading and mathematics that are now recognized as a standard and required by all 50 states in the USA [1, 32].

Two recent papers have explored—with mixed results—the relationship between narrow band mental health scores and other academic achievement tests (although not the type that is now required by the states). Breslau and his colleagues [6] conducted a secondary analysis of a longitudinal sample of 823 urban children who had been followed from birth and reported that using a domain-specific teacher report measure administered at age 6 years, internalizing, externalizing and attention problems all predicted math and reading achievement at age 17 years, although when entered in a stepwise fashion only attention problems were significant predictors. Duncan and colleagues [11], in a secondary analysis of six large longitudinal studies, found that on domain-specific behavioral measures completed by teachers or parents at school entry (ages 5–6), it was only attentional problems that predicted math and reading achievement test scores at the end of primary school (ages 11–12 years). Neither externalizing nor internalizing problems predicted academic test scores.

The relationship between broadband measures of mental health and academic outcomes has been explored in a

number of smaller studies. In a sample of 156 preschoolers who were screened with a brief teacher-rated version of the BASC (Behavior Assessment System for Children), total problem score had a strong negative correlation with reading and math achievement test scores in the first grade [24]. In samples of 166 middle school and 185 high school students, positive screening scores on the parent- or student-rated Pediatric Symptom Checklist (PSC) have been shown to be related to higher rates of school absence, poor grades and/or repeating a grade in school [14, 33].

In summary, all of the studies reviewed have significant limitations with regard to understanding the relationship between the kinds of broadband mental health screens that are recommended or required in pediatrics and the kinds of standardized achievement tests that are now the de facto standard throughout the USA. The current paper addresses these limitations by using two different broadband screens in a large sample to predict national achievement test scores much like those used in the USA. These analyses were possible because the nation of Chile employs achievement tests much like those used in the USA and because it uses broadband screens to identify students at risk in a national mental health program designed to improve educational outcomes.

Habilidades para la Vida [translated as Skills for Life (SFL)] is a school-based mental health promotion and intervention program that is offered by the National Board of School Aid and Scholarships (JUNAEB) of Chile. Beginning in the 1990s, JUNAEB started to provide programs aimed at promoting equal opportunity to education and preventing school dropout for all students in Chile. In addition to school meals, and programs correcting dental, vision, hearing and orthopedic problems that could interfere with learning, a program targeting mental health was added in 1998. SFL had evolved over the 1990s from studies of the prevalence of mental health problems in Chile, the adaptation and validation of screening instruments for that country and the development of the intervention itself. The intervention followed the three-tiered model (promotion/targeted prevention/intervention) recommended by the World Health Organization in a seminal report [19] and developed over time in collaboration with local communities and consultation from a number of authorities in children's mental health including Drs. Sheppard Kellam and Thomas Anders [10]. The program's voluntary adoption by so many local school districts is probably due to their perception that it is effective in reducing behavior problems and this in turn is one of the most important factors in its steady growth in a handful of schools in 1998 to more than 1,000 at the present time.

Skills for Life is available on a voluntary basis to selected public and subsidized private schools in Chile that meet criteria for "high risk" based in part on a formula created by the World Health Organization that takes school-level indicators of family income, maternal education and other factors into account. In Chile, about 20% of all schools are considered to be high risk by this standard. The SFL program reached national scale in 2002 with 402 schools participating, and since that time it has grown to include 1,172 schools and 181,352 students [23].

In keeping with the original WHO model, SFL offers mental health promotion activities for all students, teachers and parents in participating schools, preventive intervention workshops for children screened at risk and referral to outside professionals for students found to be at highest risk. For screening, SFL employs brief standardized measures—developed in the USA and adapted and validated in Chile—as broadband screens for mental health problems for all of the first grade students in these schools. Students identified with problems on the teacher screen are referred to a 15-session school-based preventive intervention and those with problems identified by the parent screen are referred to their primary care providers for further evaluation.

The current paper focuses on evidence on whether mental health matters in this elementary school sample: whether the mental health screening scores from first grade of more than 11,000 students predict academic achievement test scores 3 years later. If the answer is yes, then the value of broadband psychosocial screening in pediatric and educational settings would be better established and the stage would also be set for assessing the impact of large-scale interventions using mental health and academic measures like these.

Methods

Sample

All subjects in the current study began as first grade students in 2002 in schools that were participants in the Skills for Life Program. The mental health screening scores of these students from the first grade were merged with their SIMCE achievement test scores from the fourth grade in 2005.

Design, setting, and population

Data on academic testing come from the SIMCE (Sistema de Medicion de Calidad de la Education), which is required for all students in Chile in the fourth, eighth and tenth grades [31]. The analytic data set for this paper required the combination of two different databases: (1) national fourth grade SIMCE test scores including parent background data questionnaire and (2) scores from first grade teacher and parent-reported mental health screening and risk factor questionnaires. After Chilean staff had merged the data using the national student ID number as the match,

identifying information was removed and the data set was sent to the USA for the analyses reported here. This procedure was reviewed and approved as exempt by the Partners Human Research Review Board in February of 2009 as well as the requisite authorities in Chile.

Measures

Outcomes

Academic achievement For each of three subject areas (math, science and language), the SIMCE test includes 40–51 questions. In a 1999 national standardization, SIMCE scores were calculated to have a mean of 250 and a standard deviation of 50 [31]. In this paper, we report all three subtest scores, although for greater simplicity in the text, we focus primarily on the average of the three subject areas for each student.

Predictors

Mental health high risk The main predictor variables were high (vs. lower) student mental health risk as assessed independently on the Chilean versions of the Teacher Observation of Classroom Adaptation-Revised (TOCA-R) and the Pediatric Symptom Checklist (PSC). Screens were administered during the second half of the school year after teachers had known their students for at least 4 months.

The TOCA-R is a valid and reliable measure, which has been used in the USA for over two decades and was one of the primary measures in an influential series of studies of elementary school classroom behavior, intervention and adult outcomes that started in the 1990s [51]. The TOCA-R is a brief structured interview with a teacher that is administered by a trained assessor. Teachers respond to items pertaining to the child's adaptation to classroom task demands as a way to identify students with problems. The TOCA-R [26] has 31 items each scored on a six-point Likert scale from 'never' to 'always' and groups of four to seven items are summed to produce scores on six problem-specific subscales. The studies referenced above have shown that one of the TOCA-R subscales (authority acceptance) scores in elementary school is predictive of real-world health, academic and mental health problems into adulthood [25, 38]. In the 1990s, Chilean social scientists, university staff, investigators and educational officials translated, adapted and validated the TOCA-R for a Chilean context in a series of studies. This measure is called the TOCA-RR.

The SFL program uses TOCA-RR Global score in the first grade as the primary determinant of overall risk. TOCA Global Risk is a dichotomous variable assigned by the SFL national program staff at the end of the first grade on the basis of the pattern of high-risk scores on the six subscales. All

students who are at risk on the TOCA Global score in the first grade are referred to school-based workshops in the second grade. Because the cohort followed in this paper was the first in the national implementation of SFL, there were high rates (30–50%) of missing data about workshop attendance and referral to primary care, and it was not possible to study the association between the interventions and academic test scores in this cohort. However, exploratory analyses using these variables were conducted and are reported below.

The second mental health screen used in SFL is the Chilean version of the Pediatric Symptom Checklist (PSC-Cl). The PSC is a one-page questionnaire listing a broad range of children's emotional and behavioral problems [22]. Over the past two decades, the PSC has been one of the most widely used psychosocial screens for children in the USA, validated in a national sample and in many subpopulations [27] and recommended for use as a part of the Medicaid EPSDT program in many states [42]. A recent study [15] has shown a high correlation between the short version of the PSC (17) and a number of well-accepted brief measures of specific types of psychopathology (like the Children's Depression Inventory [28]) and with overall psychiatric diagnosis on the K-SADS-L [3].

The PSC also went through an extensive adaptation to the Chilean context by the same team that validated the TOCA-RR [10, 16, 17]. The PSC-Cl contains 33 items, each of which is reported by the parent as never, sometimes or often present as on the US PSC. On the PSC-Cl, the items are weighted as 1, 2 and 3, whereas on the US version they are rated as 0, 1 and 2. On both versions, a total score is obtained by summing the weighted scores for all items and higher scores indicate more problems and more serious risk. As on the US form, PSC-Cl total scores can be recoded into a dichotomous score based on a pre-determined cutoff score. For the PSC-Cl, scores of 65 or higher indicate mental health risk [10]. To obtain an estimate of overall mental health risk in the first grade, we combined the categorical TOCA-RR and PSC-Cl scores to distinguish between students who were at risk on both measures, only one or neither.

Covariates

The PSC-Cl form also assesses five risk factors: whether (1) the mother was a teenager when the child was born, (2) whether the father lives with the child, (3) whether the child lives with a relative disabled by mental illness, (4) whether the family participates in organized social activities, and (5) whether the child has a chronic illness leading to one or more school absences a month. These items are all dichotomous, requiring parents to give yes/no answers. These variables were added to the Chilean version of the PSC in 1992 and have remained a part of it ever since, because officials have found them to provide useful information [30]. The TOCA

form in both the Chilean and US versions contains an additional summary item that asks the teacher to rate the child's overall academic progress on a six-point Likert scale ranging from excellent to failing. To ascertain the impact of mental health problems on standardized test scores in the fourth grade in these analyses, we used this variable to control for academic ability in the first grade. In keeping with the dichotomous approach to risk factors shown in the analyses presented in Table 2, the Likert categories of failing, very poor and poor achievement in the first grade were grouped to operationalize a category of academic risk and contrasted with Likert ratings of fair, good and very good achievement, which were grouped together to operationalize a lack of academic risk. Scores on this one TOCA-RR item and on the five PSC-Cl risk factor items are not considered in the TOCA-RR or PSC-Cl total scores.

We also used several background factors from the SIMCE parent questionnaire as additional covariates: child gender; family socioeconomic status (SES; a recoding of the family's reported monthly income into five levels), whether the school was public or private (with or without public subsidy) and mother's and father's level of education (elementary, high school, college). These five SIMCE variables were collected during fourth grade.

Data analysis

Mental health and academic achievement

We first examined bivariate relationships between the key predictor variables and SIMCE scores without adjustments and then the same relationships taking covariates into account. We used inference via multiple imputations (MI) to make full use of observed data and to incorporate the added uncertainty due to missing data. MI is a simulation-based inferential tool operating on multiple "completed" data sets, where the missing values are replaced by real values based on random draws from their respective predictive distributions [53]. Based on this model, multiple draws were used to impute missing data due to mismatched or incomplete scores on the PSC. Our imputation model accounted for clustering due to school and relationships that are important to our substantive analyses, while also reflecting the potential causes for missing data.

Results

In 2002 when the students in the program entered school, 402 schools in 43 districts with 25,442 first grade students had elected to participate in the SFL program. Schools submitted usable SFL questionnaire data on 20,135 of these students and 17,252 (68%) of the students had completed forms for the TOCA. It was possible to match 11,185 (65%) of these students with their SIMCE scores in the fourth grade. Completed PSC-Cls were obtained for 8,510 (76%) of these children, of whom 7,903 (93%) had complete data on all five PSC-Cl risk factors and all other variables. This became the primary analytic sample for the current paper.

Table 1 compares students in this subsample ('Teacher and parent-screened sample'; column 1) with students in the 'Teacher but not parent-screened sample' (those whose

	PSC completers: PSC and TOCA in the first grade with complete SIMCE in the fourth grade[b] Teacher and parent-screened sample, $N = 7,903$	TOCA form only completers: only TOCA (no PSC) in the first grade with complete SIMCE in the fourth grade[b] Teacher but not parent-screened sample, $N = 3,282$
School financing[a]		
Public school	79%	78%
Private school/public funding	20%	22%
Private school	0%	0%
Gender[a]		
Female	50%	51%
Male	50%	49%
Socioeconomic level[a]		
Low	8%	10%
Lower middle	37%	39%
Middle	44%	41%
Upper middle	11%	11%
Upper	0%	0%**
SIMCE 3 subject summary score[b]	245.85	236.84***

Table 1 Information on analytic subsamples

** $p < .01$; *** $p < .001$

[a] Chi square tests were used to test differences in percentages for gender, financing and SES groups across the two subsamples

[b] Analysis of variance was used to compare the SIMCE three-subject summary scores for column 1 versus column 2 ($F = 85.57$, $p < .001$)

Table 2 Relationship between risk factors in the first grade and achievement test scores in the fourth grade for cases with complete data

	N (%) 7,903 (100%)	SIMCE fourth grade achievement test scores			
		Summary[a]	Math	Science	Language
Gender					
Female	3,954 (50)	244.71	238.18	245.30	250.64***
Male	3,949 (50)	246.99*	244.00***	251.39***	245.57
School financing[b]					
Public school	6,274 (79)	242.94	238.61	244.93	245.28
Private school/public funding	1,629 (21)	257.04***	250.65***	261.48***	259.01***
Socioeconomic level[b]					
Low	627 (8)	226.31	221.86	225.30	231.77
Middle low	2,918 (37)	234.72	230.69	236.57	236.90
Middle	3,458 (44)	253.91	248.73	257.19	255.80
Middle high	900 (11)	264.56***	258.85***	268.56***	262.28***
Other family/child variables					
Mother has grade school ed.	2,842 (36)	234.71	230.51	236.23	237.40
Mother has high school ed.	3,394 (43)	246.29	241.89	249.09	248.20
Mother has ed. beyond HS	1,667 (21)	263.92***	258.12***	267.47***	266.18***
Father has grade school ed.	2,856 (36)	232.80	228.40	233.55	236.45
Father has high school ed.	3,374 (43)	247.33	242.72	250.68	248.60
Father has ed. beyond HS	1,673 (21)	265.13***	259.48***	268.88***	267.03***
Child has no father at home	1,902 (24)	243.81	237.67	246.67	246.79
Lives with father in home	6,001 (76)	246.49*	242.08**	248.87	248.53
Child lives with disabled relative	888 (11)	238.58	233.88	240.47	241.40
No mentally disabled relative	7,015 (89)	246.77***	242.00***	249.34***	248.96***
Mother's age <25 years	773 (10)	240.18	235.93	242.34	242.26
Mother 25+ years old	7,130 (90)	246.46***	241.65***	248.99**	248.74**
Child has an illness	1,123 (14)	241.66	235.16	246.02	243.79
Child does not have an illness	6,780 (86)	246.54**	242.07***	248.73	248.82**
Family is socially isolated	3,897 (49)	246.00	241.21	248.38	248.40
Family is not socially isolated	4,006 (51)	245.70	240.98	248.30	247.82
Teacher-rated problem with achievement in the first grade	417 (5)	199.45	193.56	201.97	203.61
No achievement problem	7,486 (95)	248.43***	243.74***	250.97***	250.59***

Unadjusted mean scores are shown in this table

* $p < .05$; ** $p < .01$; *** $p < .001$

[a] Average of SIMCE math, language and science scores

[b] 20 subjects coded as high SES and 19 coded as private school, recoded as upper middle SES or subsidized private school because of small cell size

parents did not fill out a PSC-Cl; column 2). These two subsamples are similar overall with no differences on type of school and gender and only slight differences on socioeconomic status and summary SIMCE score, with the PSC completers significantly less likely to be from the lowest SES groups and to have higher SIMCE scores.

Table 2 shows the bivariate relationships between risk factors and standardized test scores (with no adjustments) in the fourth grade for those in the teacher- and parent-screened sample. All but one of the 11 risk factor variables were significantly associated with lower scores on SIMCE '3 subject summary score' in the fourth grade. Since the

patterns for all three subjects and summary tests were similar, in text we will describe only the SIMCE summary score findings. As shown in the first ('Summary') column under 'SIMCE fourth grade achievement test scores' in Table 2, the only exception to this pattern was with regard to the impact of gender: boys had significantly higher SIMCE summary scores (246.99) than girls (244.71) although this obscured the more complex pattern in which boys had higher science and math scores and girls had higher language test scores.

Continuing down the 'Summary' column in Table 2, public school students had scores that were about 14 points lower than students in private or subsidized private schools (242.94 vs. 257.04; $F = 116.23$, $p < .001$). Students from the lowest SES group had scores that were about 38 points lower than students from the highest (upper middle) SES group ($F = 180.85$, $p < .001$). Students whose mothers had only grade school educations scored an average of 29 points lower than students whose mothers had educations beyond high school ($F = 210.44$, $p < .001$). Students whose fathers had only grade school educations scored an average of 32 points lower than students whose fathers had education beyond high school ($F = 265.08$, $p < .001$). Students who had had teenaged mothers, absent fathers, family histories of mental illness or frequent school absences had SIMCE scores that were from three to eight points lower than did students without these risk factors. The final rows of Table 2 show that the strongest predictor of standardized test scores in the fourth grade is teacher rating of academic achievement problems in the first grade. For the 5% or so of students who were rated by teachers as having poor, very poor or failing achievement in first grade, the average fourth grade SIMCE scores were 49 points (almost a full standard deviation) lower than they were for students rated by their teachers as having fair,

good or very good academic achievement in the first grade (199.45 vs. 248.43; $F = 445.83$, $p < .001$).

Table 3 shows the relationship between mental health risk assessed in the first grade (as measured by the TOCA-RR and PSC-Cl separately and combined) and standardized test scores in the fourth grade, again with no adjustment for covariates. In the analytic sample, 11% of the students were classified as being at risk on the TOCA-RR, as were 11% of the students on the PSC-Cl. On the two measures combined, 3% of students were at risk on both measures, 16% on one and 81% on neither. Students rated as being at risk on the TOCA-RR in the first grade had fourth grade SIMCE summary scores that were about 28 points lower than students not at risk on Global TOCA-RR score in the first grade (220.95 vs. 248.98; $F = 284.61$, $p < .001$). Students rated as being at risk on the PSC-Cl in the first grade had fourth grade SIMCE summary scores that were 22 points lower than students not at risk on PSC-Cl in the first grade ($F = 176.51$, $p < .001$). As shown in the last three rows in Table 3, students with both teacher- and parent-rated mental health problems in the first grade scored an average of nearly 40 points lower on summary SIMCE score than students coded as being at risk on neither of the screens ($F = 198.89$, $p < .001$). Students coded as being at risk on only one of the two screens had SIMCE scores in between.

Table 4 shows the same relationships as Table 3, except that these results control for all of the covariates listed in Table 2 and use linear mixed-effects models to control for school-level variation and to handle missing data in the covariates and the responses using inference by multiple imputation. Even after these adjustments, teacher- and parent-rated mental health risk assessed separately and/or together in the first grade were still found to be significantly associated with poorer standardized test scores in the

Table 3 Relationship between mental health and behavior risk in the first grade and academic test scores in the fourth grades unadjusted for covariates

Risk in the first grade	N (%) 7,903 (100%)	SIMCE academic achievement test score in the fourth grade			
		Summary SIMCE	Math	Science	Language
TOCA-RR risk					
Behavior risk	884 (11)	220.95	214.35	224.17	224.32
Behavior all right	7,019 (89)	248.98***	244.46***	251.38**	251.10***
PSC-Cl risk					
Mental health risk	870 (11)	225.93	220.05	228.67	229.07
Mentally healthy	7,033 (89)	248.31***	243.69***	250.77***	250.46***
TOCA-RR and PSC-Cl risk					
Behavior problem on both	233 (3)	211.29	206.24	213.01	214.62
Behavior problem on one	1,288 (16)	227.81	221.13	231.25	231.04
Behavior all right	6,382 (81)	250.75***	246.39***	253.08***	252.78***

** $p < .01$; *** $p < .001$

Table 4 Relationship between mental health and behavior risk in the first grade and academic test scores in the fourth grades adjusted for covariates, school-level clustering, and missing data uncertainty via multiple imputation

Risk in the first grade	SIMCE academic achievement test score in the fourth grade			
	Summary SIMCE	Math	Science	Language
TOCA-RR Risk				
Behavior risk	221.83	218.31	226.25	223.17
Behavior all right	242.35***	236.10***	243.70***	244.21
PSC-Cl Risk				
Mental health risk	225.56	219.31	225.69	230.99
Mentally healthy	241.56***	237.22***	242.26***	241.40***
TOCA-RR and PSC-Cl Risk				
Behavior problem on both	218.32	215.89	224.75	225.09
Behavior problem on one	230.98	224.40	234.15	234.55
Behavior all right	251.00***	242.39***	249.16***	249.56***

Gender, school, SES, mother and father education, father at home, family member sick, teenage mother, child sick often, family social isolation, teacher-rated academic problem in the first grade

*** $p < .001$

fourth grade. In these analyses, the apparent difference in SIMCE average score related to mental health risk decreases slightly from 28 to 21 points on the TOCA-RR, from 22 to 16 points on the PSC-Cl, and from 39 to 33 points on the two measures combined, still about two-thirds of a standard deviation in summary SIMCE score. The pattern on all of the SIMCE subtests is the same.

In a final set of analyses not shown here, the data on workshop attendance and referral to specialists were added as covariates to the multivariate equations shown in Table 4, and the results did not change appreciably. The relationship between TOCA-RR and PSC-Cl risk separately and combined and SIMCE summary score were still significant and of the same magnitude as reported above.

Discussion

Results from this study demonstrate that being identified with a mental health problem in the first grade on a broad-band teacher, parent or combined screen predicted significantly poorer performance 3 years later on routinely given standardized achievement tests, the most widely used benchmark for children's real-world functioning in the USA and in many other countries. The study also shows that mental health problems stand on their own as an independent risk factor after controlling for other major risk factors.

This study allowed us to quantify the impact of mental health problems on student achievement. Using a multi-variable analysis of variance approach, when the two measures of mental health were combined, the unadjusted 39-point differential in achievement test scores between students with both and students with neither of the mental

health screening risks was slightly larger than the 38-point difference in achievement for the poorest versus upper middle class children, making mental health the second most powerful predictor of achievement in the study (after the 50-point impact of teacher rating of academic performance at baseline). Paternal and maternal education were the fourth and fifth most powerful variables and of almost comparable magnitude (impacts of 32 and 29 points, respectively), but the other social risk factors studied were of a smaller magnitude (3–8 points), although still also significantly associated with lower test scores. Since socioeconomic status, native ability and parental education [9, 43] have been long established as the most powerful predictors of academic achievement in elementary school, this study's finding that mental health is of comparable magnitude is of great importance since it provides empirical support for efforts in Chile, the USA and international agencies such as the WHO that are based on the premise that mental health really does matter for children.

Our findings are limited to some extent by the large amount of missing data on some of the variables in this real-world sample. However, when multiple imputation, a procedure that can be used to adjust the data in such a way as to control for the 'missingness' of some of the variables was used in the current analyses, the observed relationships between mental health screening scores and achievement test scores were still significant and of almost the same magnitude.

Another limitation of the current study is that it does not address the key question of the impact on achievement of the intervention that was provided to some of the students in this sample. With about one-third of the analytic sample missing, follow-up data on attendance at the school-based

workshops and/or referral to outside specialists in this first year of national data collection, it was not possible to assess the impact of interventions in this sample. A paper in preparation [34] based on a later cohort, which has less missing data on these variables, does report a significant relationship between participating in the SFL workshops and improved academic outcomes.

It should also be noted that this study was done in a single country. Chile is in some ways unique in its combination of characteristics of developed and developing nations [8]. As one of the most prosperous nations in the southern hemisphere, it has resources for screening and intervention that other developing nations may lack and it has the flexibility to design large new programs that many developed countries may lack. Overall, however, the prevalence of psychosocial problems and the need for academic achievement are concerns for children around the world.

A final limitation of this study is that although the SIMCE test is thought to be similar to standardized academic achievement tests given in the USA, there have been few if any studies that prove this. Therefore, the power of the PSC and TOCA to predict achievement test scores in the USA remains to be demonstrated.

Conclusion

Although the limitations noted above constrain its conclusions somewhat, the current study nevertheless provides the strongest evidence to date that students with overall, broadband mental health risk have lower levels of subsequent academic achievement as measured by standardized academic achievement tests when compared with students who are not at overall mental health risk. Unlike poverty, parental education and preexisting academic ability—the other major predictors of academic success in this study—mental health is a risk factor that may yield to intervention.

In the USA, programs such as Medicaid and Head Start have long acknowledged, but never fully acted upon, the need to address the mental health needs of children. Neither program has effectively implemented mental health screening or treatment on a national scale despite program standards that require this [21, 42]. In the rest of the world too, the burden of mental health problems has long been acknowledged, but until now not specifically addressed through national programs [12, 19, 52]. By providing data on the strong association between psychosocial risk identified through screening and poorer achievement test scores in a large real-world sample, the current study provides clear evidence that mental health matters for children. We hope that this evidence will encourage efforts to fully implement the recommended mental health standards for

screening and intervention in pediatric and educational settings in the USA and other countries as well. As recommendations for screening are followed increasingly throughout the world, it should be possible to ascertain whether current findings about the relationship between mental health measures and academic achievement test scores are replicated in other countries, which use different tests. Perhaps even more importantly, existing data in Chile can be analyzed to ascertain whether a relatively inexpensive school-based intervention is associated with improved outcomes. If preliminary findings that show improved outcomes in Chile are confirmed, then child advocates in other countries will have the evidence they need to create and test similar programs in their own countries.

Conflict of interest All of the Chilean authors received salary support from the Chilean government for administering the mental health program. All US authors except Dr. Jellinek received salary or consultation support for data analysis or report writing. Drs. Jellinek and Murphy are the authors of the PSC, but they received no income from this since it is available free in the public domain. No other authors reported any financial disclosures or conflict of interest.

References

1. 107th Congress (enacted January 8, 2002) No Child Left Behind Act of 2001. In: PUBLIC LAW 107–110, JAN 8 2002 115 STAT. 1425
2. Aluja A, Blanch A (2002) The children depression inventory as predictor of social and scholastic competence. Eur J Psychol Assess 18:15
3. Ambrosini PJ (2000) Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). J Am Acad Child Adolesc Psychiatry 39:49–58
4. American Academy of Pediatrics (2004) Policy statement: school-based mental health services. Pediatrics 113:1839–1845
5. American Academy of Pediatrics Task Force on Mental Health (2010) Appendix S4: the case for routine mental health screening. Pediatrics 125:S133–S139
6. Breslau J, Miller E, Breslau N, Bohnert K, Lucia V, Schweitzer J (2009) The impact of early behavior disturbances on academic achievement in high school. Pediatrics 123:1472–1476
7. Cancecko-Llego CD, Castillo-Carandang NT, Reyes AL (2009) Validation of the pictorial pediatric symptom checklist-filipino version for the psychosocial screening of children in a low-income urban community. Acta Medica Philippina 43:62–68
8. Contreras D, Bustos S, Sep'ulveda P (2007) When schools are the ones that choose: the effect of screening in Chile. In: Serie Documentos de Trabajo. Departamento de Economía, Universidad de Chile

9. Davis-Kean P (2005) The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. J Fam Psychol 19:294–304

10. de la Barra F, Toledo V, Rodriguez J (2005) Prediction of behavioral problems in Chilean school children. Child Psychiatry Hum Dev 35:227–243

11. Duncan G, Dowsett C, Claessens A, Magnuson K, Huston A, Klebanov P, Pagani L, Feinstein L, Engel M, Brooks-Gunn J, Sexton H, Duckworth K, Japel C (2007) School readiness and later achievement. Dev Psychol 43:1428–1446

12. Fisher JRW, Cabral de Mello M, Izutsu T, Vijayakumar L, Belfer M, Omigbodun O (2011) Adolescent mental health in resource-constrained settings: a review of the evidence of the nature, prevalence and determinants of common mental health problems and their management in primary health care. J Soc Psychiatry 57(1 Suppl):v–vii, 9–116

13. Fröjd S, Nissinena E, Pelkonenb MU, Marttunen MJ, Koivistoa A, Kaltiala-Heinoa R (2008) Depression and school performance in middle adolescent boys and girls. J Adolesc 31:14

14. Gall G, Pagano ME, Desmond MS, Perrin JM, Murphy JM (2000) Utility of psychosocial screening at a school-based health center. J Sch Health 70:292–298

15. Gardner W, Lucas A, Kolko D, Campo J (2007) Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. J Am Acad Child Adolesc Psychiatry 46:611–618

16. George M, Siraqyan X, Mores R, de la Barra F, Rodríguez J, López C, Toledo V (1995) Adaptation and validation of two screening instruments of mental health of first grade schoolchildren of Santiago. Revista de Psicologıa 5:17–25

17. George M, Squicciarini A, Zapata R, Guzmán MP, Hartley M, Silva C (2004) Detection of early mental health risk factors in primary schools. Revista de Psicologia 13:9–20

18. Goldberg C (2007) Mental screening for young to begin: mass. doctors to offer questionnaires for children on Medicaid. The Boston Globe. NY Times Co., Boston

19. Hendren R, Birrell Weisen R, Orley J (1994) Mental health programmes in schools. World Health Organization, Geneva

20. Ishizaki Y, Fukai Y, Kobayashi Y (1997) Preliminary study of validation of the Japanese version of the pediatric symptom checklist by Jellinek of Massachusetts General Hospital- a new screening method for children with psychosocial and psychosomatic dysfunction. J Jpn Pediatr Soc 101:1679–1985

21. Jellinek MS, Bishop Joseph SJ, Murphy JM, Zigler E (2005) Mental health in head start; leave no child behind. NHSA DIA-LOG 8:25–35

22. Jellinek MS, Murphy JM, Burns BJ (1986) Brief psychosocial screening in outpatient pediatric practice. J Pediatr 109:371–378

23. JUNAEB Chilean Department of Education (2009) Skills for life program; Report on the 2008 school year. JUNAEB, Santiago

24. Kamphaus RW, Thorpe JS, Winsor AP, Kroncke AP, Dowdy ET, VanDeventer MC (2007) Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. Educ Psychol Measur 67:1–15

25. Kellam Sheppard G, Brown H, Poduska JM, Ialongo NS, Wang W, Toyinbo P, Petras H, Ford C, Windham A, Wilcox HC (2008) Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. Drug Alcohol Dependence 95(Suppl 1):S5–S28

26. Kellam SG, Rebok GW, Mayer LS, Ialongo N (1994) Depressive symptoms over first grade and their response to a developmental epidemiologically based preventive trial aimed at improving achievement. Dev Psychopathol 6:463–481

27. Kelleher K, McInerny T, Gardner W, Childs G, Wasserman R (2000) Increasing identification of psychosocial problems: 1979–1996. Pediatrics 105:1313–1321

28. Kovacs M (1985) The children's depression, inventory (CDI). Psychopharmacol Bull 21:995–998

29. Kuhlthau K, Jellinek M, White G, VanCleave J, Jack Simons J, Michael Murphy JM (2011) Increases in behavioral health screening in pediatric care for Massachusetts medicaid patients. Arch Pediatr Adolesc Med. doi:10.1001/archpediatrics.2011.18

30. López C, de la Barra F, George M, Toledo V, Rodríguez J (1996) Características de la Familia y Conducta de los Niños Evaluada por sus Padres y Profesores. Revista de Psiquiatria 12:195–1999

31. Meckes L, Carrasco R (2010) Two decades of SIMCE: An overview of the National Assessment System in Chile. Assessment in Education: Principles, Policy and Practice 17:233–248

32. Meier D, Wood G (2004) Many children left behind: how the No Child Left Behind Act is damaging our children and our schools. Beacon Press, Boston

33. Murphy JM, Jellinek M, Milinsky S (1989) The pediatric symptom checklist: validation in the real world of middle school. J Pediatr Psychol 14:629–639

34. Murphy JM, Masek B, Babcock R, Jellinek M, Gold J, Drubner S, Sklar K, Hacker K (2011) Measuring outcomes in outpatient child psychiatry: the contribution of electronic technologies and parent report. Clin Child Psychol Psychiatry 16:146–160

35. Murray CJ, Lopez AD (1996) Evidence-based health policy–lessons from the global burden of disease study. Science 274:740–743

36. Muzzolon SRB (2008) Evaluation of the use of the Pediatric Symptom Checklist as screening for emotional and psychosocial problems in a public school in Curitiba. Brazil Arquivos de Neuro-Psiquiatria 66:604

37. National Research Council and Institute of Medicine (2009) Preventing mental, emotional, and behavioral disorders among young people: progress and possibilities. Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth, and Young Adults: research advances and promising interventions. In: O'Connell ME, Boat T, Warner KE (eds) Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC

38. Petras H, Chilcoat H, Leaf P, Ialongo N, Kellam S (2004) Utility of TOCA-R scores during the elementary school years in identifying later violence among adolescent males. J Am Acad Child Adolesc Psychiatry 43:88–96

39. Petras H, Ialongo N, Lambert S, Barrueco S, Schaeffer C, Chilcoat H, Kellam S (2005) The utility of elementary school TOCA-R scores in identifying later criminal court violence among adolescent females. J Am Acad Child Adolesc Psychiatry 44:790–797

40. Ponsor MA (2006) ROSIE D. et al. v. Mitt ROMNEY et al. In: United States District Court District of Massachusetts (ed) Civil Action No. 01 30199 MAP. Boston, MA, p 410 F.Supp.412d 418

41. Reijneveld SA, Vogels AG, Hoekstra F, Crone MR (2006) Use of the pediatric symptom checklist for the detection of psychosocial problems in preventive child healthcare. BMC Public Health 6:197–204

42. Semansky RM, Koyanagi C, Vandivort-Warren R (2003) Behavioral health screening policies in medicaid programs nationwide. Psychiatr Serv 54:736–739

43. Siran S (2005) Status and academic achievement: a meta-analytic review of research. Rev Edu Res 75:417–453

44. Thun-Hohenstein L, Herzog S (2008) The predictive value of the pediatric symptom checklist in 5-year-old Austrian children. Eur J Pediatr 167:323–329

45. US Department of Health and Human Services (2000) Healthy People 2010, 2nd ed. With understanding and improving health and objectives for improving health, 2 vols. In: US Department of Health and Human Services Office of Disease Prevention and Health Promotion, Washington, DC

46. US Department of Health and Human Services (1999) Mental health: a report of the surgeon general. In: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health, Rockville, MD

47. US Preventive Services Task Force (2009) Screening and treatment for major depressive disorder in children and adolescents: US Preventive Services Task Force recommendation statement. Pediatrics 123:1611

48. Üstün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL (2004) Global burden of depressive disorders in the year 2000. Br J Psychiatry 184:386–392

49. Wagner M (1995) Outcomes for youths with serious emotional disturbances in secondary school and early adulthood. Future Child 5:90–112

50. Warnick EM, Bracken MB, Kasl S (2008) Screening efficiency of the Child Behavior Checklist and Strengths and Difficulties Questionnaire: a systematic review. J Child Adolesc Ment Health 13:140–147

51. Werthamer-Larsson L, Kellam SG, Wheeler L (1991) Effects of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. Am J Community Psychol 19:585–602

52. WHO (2003) Caring for children and adolescents with mental disorders; setting WHO directions. In: Meeting on Caring for Children and Adolescents with Mental Disorders. Geneva, Switzerland

53. Yucel RM, Demirtas H (2010) Impact of non-normal random effects on inference by multiple imputation: a simulation assessment. Comput Stat Data Anal 54:790–801