


RESEARCH

Open Access



Learn Kind: a cluster randomized controlled trial

Mehmet Necip Tunç^{1,2}, Chloe San Miguel¹, Megan Peterson¹, Rebecca Reed¹, Jaclyn Lindsey¹ and Oliver Scott Curry^{1,3*} 

*Correspondence:

Oliver Scott Curry

oliver.curry@anthro.ox.ac.uk

¹KindLab, Kindness.org, New York, USA

²Department of Social Psychology, Tilburg University, Tilburg, Netherlands

³School of Anthropology and Museum Ethnography, University of Oxford, Oxford, UK

Abstract

Can you teach schoolchildren to be kinder? We conducted a randomized controlled trial of Learn Kind—a kindness-based SEL curriculum for grades K-8, designed by kindness.org. Teachers—randomly assigned to either a treatment group, or a waitlist group—completed measures and questions (kindness grades, happiness, the Strengths and Difficulties Questionnaire, math and reading grades) reporting on approximately 1202 students at three time points (Time 1–Time 3). We found, first: from Time 1–2, the treatment group significantly improved on all measures ($\bar{d}=0.38$); the waitlist group did not significantly improve on any measure ($\bar{d}=0.03$); and the treatment group improved on all measures significantly more than the waitlist group ($\bar{d}=0.34$). Second, the improvements in the treatment group on all measures (except 'attention') persisted from Time 2–3 ($\bar{d}=0.01$); and on all measures the treatment group had improved Time 1–3 ($\bar{d}=0.38$). Third, from Time 2–3 the waitlist group improved significantly on happiness, emotions, and reading ($\bar{d}=0.09$), but improved significantly more than the treatment group only on emotions and attention ($\bar{d}=0.08$). Taken together, the study found that Learn Kind had a small positive effect on kindness, happiness, strengths and difficulties, and academic performance. The implications, limitations, and future directions of this research are discussed.

Keywords Social and emotional learning, Randomized control trial, Kindness, Happiness

Kindness is typically understood as *actions* intended to *benefit* others, at some *cost* to the actor—an ‘ABC’ model of kindness [1, 7, 19]. Whereas it was once a problem to explain why individuals would ever be kind to one another, there now exist numerous theories of human social, cooperative and altruistic behaviour. These theories—kin altruism, mutualism, reciprocal altruism, and competitive altruism—make it possible to explain a variety of different types of kindness (for example, love, sympathy, gratitude and heroism) [7]. And they predict that people will help family, friends, community members, spouses, and even strangers under some conditions. Previous empirical work has documented the degree to which individuals—including children—are willing to pay a cost to benefit others [8, 9, 12]. Research in developmental psychology has found that kindness-related traits emerge early in infancy [10], and grow gradually through childhood [2, 6, 17, 18].



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Research has also found that individual differences in kindness (and related traits like altruism, benevolence and compassion) are partly due to differences in genes, but largely due to differences in non-shared environmental experiences [22]. Schools are one possible source of these experiences.

Previous meta-analyses of social and emotional learning (SEL) programs in schools have shown immediate effects on a variety of outcomes. For example: Durlak [11] found positive effects of SEL on: positive social behaviour ($g = 0.23$), SEL skills ($g = 0.57$), emotional distress ($g = -0.24$), and academic performance ($g = 0.27$); Sklad [21] found positive effects of SEL on prosocial behaviour ($d = 0.39$), social skills ($d = 0.70$), and academic performance ($d = 0.46$); Corcoran [5] found positive effects of SEL on reading ($ES = 0.25$), mathematics ($ES = 0.26$), and science ($ES = 0.19$); and Cipriano [4] found positive effects of SEL on school climate/safety ($g = 0.293$), civic attitudes/behaviours ($g = 0.255$), SEL skills ($g = 0.219$), peer relationships ($g = 0.222$), attitudes/beliefs ($g = 0.209$), prosocial behaviours ($g = 0.178$), externalizing behaviours ($g = 0.162$), emotional distress ($g = 0.140$), school functioning ($g = 0.122$), and academic achievement ($g = 0.112$). These effects can persist for months to years following the intervention [4, 21, 23].

Here we conduct a randomized controlled trial of a new SEL program, Learn Kind—a fourteen-part kindness-based SEL curriculum for grades K-8, designed by kindness.org.¹

1 The Learn Kind Curriculum

The development of Learn Kind began with an extensive review of available kindness programs and interventions, and previous studies of their efficacy, and of prosocial behaviour, child development, and kindness in classrooms more broadly (see SM section “Learn Kind Curriculum Research & Bibliography” for an extended bibliography). We also conducted a survey of K-8 U.S. educators to identify their unmet needs. Drawing on this research, and using the Understanding by Design framework [25], we developed Learn Kind—a modular curriculum that frames kindness within the context of scientific inquiry. In particular, the curriculum included an experiential component, in which students were invited to perform an act of kindness and to record its effects on their own social and emotional well-being. Thus Learn Kind invites students to discover both the skills needed to increase kindness, and the impact it can have when they do. The end result was a Learn Kind curriculum consisting of eight teacher modules and six student modules (14 in total), available in three versions for grades K-2, 3–5, and 6–8. Learn Kind was delivered as a resource bundle including scaffolded lesson plans, strategies for teacher support, social and emotional skill development, and classroom community development. The curriculum included instructional support for in-person, online, and hybrid learning environments.²

In summary, Learn Kind provides an English language, short duration (\leq one semester); standalone, self-contained resource (that does not require teacher training or licensing), for K-8 students, and features an experiential component. A review of the 214 named SEL programs on which there have been published studies reveals that Learn Kind is the only SEL program to meet these criteria [4].

¹In the U.S., grades K-8 generally include children ages 5-6 in Kindergarten, 6-7 in 1st grade, 7-8 in 2nd grade, 8-9 in 3rd grade, 9-10 in 4th grade, 10-11 in 5th grade, 11-12 in 6th grade, 12-13 in 7th grade, and 13-14 in 8th grade.

²An overview of the curriculum is available in the Supplementary Materials section “Overview of the Learn Kind curriculum”; more information, and access to the materials, is available here: <https://kindness.org/learnkind>.

Learn Kind was piloted in 2020–21 and 2021–22 school years. Pilot studies conducted each year provide some evidence that Learn Kind is effective in increasing kindness. Teachers (2020–21, $n=58$; 2021–22, $n=30$) completed pre- and post surveys in which they awarded their students 'kindness grades' (A, B, C, D, F). Kindness grades were significantly higher after teaching the course ($d=0.79$, $p<0.001$; $d=0.47$, $p=0.016$)—equivalent to an increase in the median grade from 'C' to 'B'. However, the lack of a control group in these pre-post observational studies means that it is not possible to draw any definitive conclusions about the causal effects of Learn Kind on kindness.

To overcome this limitation, in the school year 2022–23, we conducted a cluster-randomized waitlist-controlled trial of the Learn Kind program. Our general research question was whether Learn Kind would perform as well as previous SEL programs (as documented by previous meta-analyses). More specifically, our hypothesis was that students in the Learn Kind treatment condition (Time 1–Time 2) would show a greater improvement in the range of outcome variables typically used to assess SEL programs (kindness, happiness, social and emotional problems, and academic performance) than students in the control condition, and that these effects would persist for at least the duration of the study (that is, for a further 3 months, from Time 2–Time 3).

2 Method

In the 2022–23 school year, teachers delivering the Learn Kind curriculum in the US were invited by email to take part in the randomized control trial.³ Out of the 1791 teachers contacted, 970 began the initial interest survey, with 676 completing it. From there, 143 people started the registration process, and 95 completed their registration. A total of 60 participants began the randomized controlled trial (RCT), with 49 successfully completing it.

Participating teachers were required to answer a series of questions about each of their students at three time points: in September 2022 before the trial started (T1, baseline) to establish pre-intervention measures; in December 2022 after the fall semester (T2) to evaluate the initial impact of the program on the Learn Kind treatment group (compared to the waitlist control group); and in April 2023 after the spring semester (T3) to assess whether any changes in the Learn Kind treatment group had persisted, and to investigate changes in the waitlist group.

In each survey, for each of their students, teachers were asked: "What overall 'kindness grade' would you give to this student?" (A, B, C, D, F); and, "How happy is the student?" (from 1-Extremely Unhappy to 5-Extremely Happy). Teachers completed a short 15-item version of the Strengths and Difficulties Questionnaire [13, 14], comprising the following five subscales: emotions, attention, behaviour, peer relations, and prosociality.⁴ Teachers were also asked "What is the student's most recent report card grade in [math and reading?]" (A, B, C, D, F). At T1, we also measured teaching experience (in years), teacher level of education, teacher and student gender, classroom grade level, number

³We conducted a power analysis to determine the necessary sample size (for clusters/teachers and students) we need to have in order to reach .80 power with our planned analysis. We expected up to 30% dropout rates in each group. For a multilevel model with 3 time points and with pre-treatment subject $icc=0.25$, $icc\ slope=0.05$, variance ratio = 0.02, and Cohen's $d=0.23$, our power analysis indicated 29 teachers per experimental group (58 classrooms in total) and 24 students per teacher as the required sample size.

⁴These scales have been coded such that higher scores indicate positive outcomes, that is, fewer problems. See SM section "SDQ Items" for items.

of students in the class, and name of the school. The surveys were hosted on Qualtrics. Teachers were paid \$500 for completing the surveys at T1 and T2, and were paid an additional \$250 for completing the survey at T3 (for a total of \$750).

Before each survey, participating teachers provided informed consent. Note that the data was not collected from students directly. Teachers reported on students using anonymized IDs unavailable to the researchers. As the primary goal was evaluation of the Learn Kind curriculum, Ethical and Independent Review Services (eandireview.com) determined the project did not require IRB review using their *Determination of Research with Human Subjects* assessment process. The consent form and ethics determination are available on in the 'Materials' files on OSF: https://osf.io/e25p7/?view_only=c3b351f5764b42a7a626a1077537be88. All methods were carried out in accordance with relevant guidelines and regulations.

We also collected archival data for school SES (median income in the school district according to the US census data [24], and school quality [15], and we recorded the Learn Kind online platform user interaction data (download history).

After completing the T1 survey, teachers were randomly assigned to an experimental condition: treatment, or waitlist (see Fig. 1).⁵ The treatment group taught Learn Kind in the fall of 2022; the waitlist group taught Learn Kind in the spring of 2023. This cross-over design allows us to: (1) compare the results of the treatment and waitlist groups T1-2; (2) investigate whether any changes in the treatment group persisted T2-3; and (3) investigate the pre-post results of the waitlist group T2-3.⁶

3 Results

60 teachers, reporting on 1414 students (50% female), completed surveys at T1.⁷ Further demographic details of the sample are available in Table 1.⁸ Descriptives for the main dependent variables, by time point and condition, are shown in Table 2, and Fig. 2.

We analyzed the data using multilevel models, with student data nested within teachers. For each outcome, the models included Condition (treatment vs waitlist), Time (T1 vs T2 vs T3), and their interaction as fixed effects, and random intercepts for student

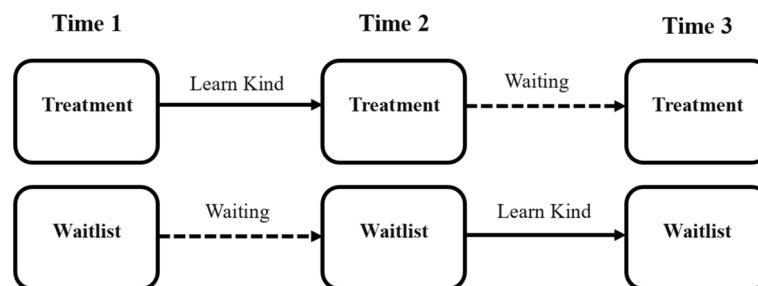


Fig. 1 Schematic of the cross-over design

⁵ Participating teachers were randomly allocated to treatment and waitlist groups using the `complete_ra` function in the R package `randomizr`.

⁶ Materials, data, and pre-registered hypotheses and analysis are available on OSF: https://osf.io/e25p7/?view_only=c3b351f5764b42a7a626a1077537be88.

⁷ When creating row-level data that contained unique student-teacher combinations, we noticed a small number (66 of 1472) of duplicate rows that may have arisen due to teacher-related errors (for example, by entering the same student ID twice). For more details, see the Supplementary Material section: "Resolving student-teacher discrepancies".

⁸ A comparison between our sample's demographics and national demographics is provided in Appendix D of the Supplemental Materials.

Table 1 Demographics

	Category	N	%	Mean	SD
Gender (Student)	Female	694	50%		
	Male	698	50%		
Gender (Teacher)	Female	53	88%		
	Male	7	12%		
Grade (Student)	K	45	3%		
	1	94	7%		
	2	151	11%		
	3	157	11%		
	4	206	15%		
	5	254	18%		
	6	171	12%		
	7	160	11%		
	8	176	12%		
Class Size		60		28.02	23.72
Median Income (by school ZIP Code)		45		80,308.31	30,827.1
Quality of Education (by school)		42		5.71	1.73
Education (Teacher)	Bachelors	22	37%		
	Masters	37	62%		
	Doctorate	1	2%		
Years of Experience (Teacher)		60		12.52	8.07

and teacher. Because these models estimate and compare slopes—that is, the changes between timepoints—they account for differences in baselines. Due to the large number of comparisons, we set the alpha at $p = 0.005$. A summary of the results is presented in Table 3. (Full analysis and models are available in file ‘2-0_LK_data_long.omv’ on OSF: https://osf.io/e25p7/?view_only=c3b351f5764b42a7a626a1077537be88).

First, we investigated the results of the treatment and waitlist groups T1-2. Comparing the pre-post results, we found that: the treatment group significantly improved on all measures (Column 1a: $\bar{d} = 0.38$); whereas the waitlist group did not significantly improve on any measure (Column 1b: $\bar{d} = 0.03$). We then investigated whether the treatment group had improved more than the waitlist control group, by examining the interaction between condition and time. We found that the treatment group improved on all measures significantly more than the waitlist group (Column 1c: kindness, $d = 0.49$; happiness, $d = 0.36$; emotions, $d = 0.37$; behaviour, $d = 0.36$; attention, $d = 0.48$; peers, $d = 0.34$; prosociality, $d = 0.37$; maths, $d = 0.20$; reading, $d = 0.13$; $\bar{d} = 0.34$.)

Second, we investigated whether the improvements in the treatment group persisted T2-3. Comparing the pre-post results, we found: that improvements in all measures persisted, or improved further (except attention) T2-3 (Column 2a: $\bar{d} = 0.01$); and on all measures the treatment group had improved from T1-3 (Column 2b: kindness, $d = 0.48$; happiness, $d = 0.45$; emotions, $d = 0.33$; behaviour, $d = 0.32$; attention, $d = 0.30$; peers, $d = 0.48$; prosociality, $d = 0.32$; maths, $d = 0.36$; reading, $d = 0.39$; $\bar{d} = 0.38$).

Third, we investigated the results of the waitlist group T2-3. Comparing the pre-post results, we found that the waitlist group improved significantly only on happiness, emotions, and reading ($\bar{d} = 0.09$). Strictly speaking, these pre-post results are the only ones we can report, because there is no genuine untreated control group in this part of the study (T2-3). However, for the sake of completeness we also investigated whether the waitlist group had improved more than the treatment group T2-3, by examining the interaction between condition and time. We found that the waitlist group had improved

Table 2 Descriptives

	Time	Condition	Teacher N	Student N	Student Mean	Student SD
Kindness	1	Waitlist	31	665	4.07	0.89
		Treatment	29	749	3.78	1.09
	2	Waitlist	25	548	4.14	0.91
		Treatment	24	624	4.29	0.87
	3	Waitlist	24	509	4.18	0.81
		Treatment	20	520	4.16	0.84
Happiness	1	Waitlist	31	665	3.92	1.03
		Treatment	29	749	3.79	1.06
	2	Waitlist	25	548	4.01	0.97
		Treatment	24	624	4.21	0.92
	3	Waitlist	24	509	4.20	0.82
		Treatment	20	520	4.23	0.82
Emotions	1	Waitlist	31	665	2.56	0.55
		Treatment	29	749	2.48	0.60
	2	Waitlist	25	548	2.55	0.56
		Treatment	24	623	2.65	0.50
	3	Waitlist	24	509	2.68	0.45
		Treatment	20	520	2.64	0.50
Behaviour	1	Waitlist	31	665	2.70	0.45
		Treatment	29	749	2.56	0.55
	2	Waitlist	25	548	2.68	0.45
		Treatment	24	624	2.72	0.45
	3	Waitlist	24	509	2.70	0.41
		Treatment	20	520	2.70	0.44
Attention	1	Waitlist	31	665	2.50	0.61
		Treatment	29	749	2.31	0.72
	2	Waitlist	25	548	2.46	0.61
		Treatment	24	623	2.60	0.58
	3	Waitlist	24	509	2.50	0.57
		Treatment	20	520	2.52	0.59
Peer relations	1	Waitlist	31	665	2.62	0.45
		Treatment	29	748	2.49	0.52
	2	Waitlist	25	548	2.65	0.42
		Treatment	24	624	2.67	0.41
	3	Waitlist	24	509	2.64	0.41
		Treatment	20	520	2.70	0.40
Prosociality	1	Waitlist	31	665	2.46	0.58
		Treatment	29	744	2.36	0.64
	2	Waitlist	25	548	2.49	0.55
		Treatment	24	624	2.60	0.53
	3	Waitlist	24	509	2.54	0.51
		Treatment	20	520	2.56	0.53
Math	1	Waitlist	30	634	3.90	0.98
		Treatment	29	749	3.74	1.08
	2	Waitlist	25	548	3.95	1.07
		Treatment	24	624	3.97	1.02
	3	Waitlist	22	492	4.03	0.96
		Treatment	20	520	4.06	0.91
Reading	1	Waitlist	30	634	3.89	0.97
		Treatment	29	748	3.75	1.06
	2	Waitlist	25	548	3.98	1.05
		Treatment	24	624	4.00	0.95

Table 2 (continued)

	Time	Condition	Teacher N	Student N	Student Mean	Student SD
	3	Waitlist	23	508	4.12	0.96
		Treatment	20	520	4.13	0.88
Mean per condition			25.41	600.85		
Mean students/teacher per condition				23.65		

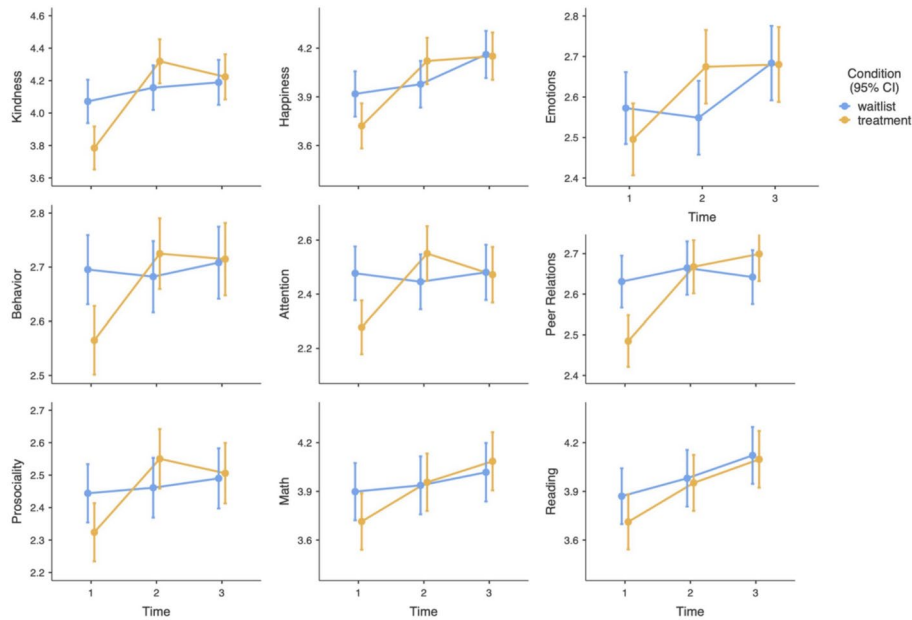


Fig. 2 Visualisation of mixed model results, regressing outcomes onto groups and time

significantly more than the treatment group only on emotions and attention (Column 3b: kindness, $d = 0.14$; happiness, $d = 0.16$; emotions, $d = 0.24$; behaviour, $d = 0.08$; attention, $d = 0.18$; peers, $d = -0.12$; prosociality, $d = 0.13$; maths, $d = -0.05$; reading, $d = 0.00$; $\bar{d} = 0.08$).⁹

We also ran a series of models to control for the effects of, and test for moderation by, teaching experience (in years), teacher level of education, teacher and student gender, classroom grade level, number of students in the class, school SES (median income in the school district according to the US census data), school educational quality (Great Schools ratings), and the total amount of materials the teachers downloaded from the Learn Kind website. These control and moderation models made no substantive difference to the results (see file ‘2-2_LK_control_intercepts.omv’ and ‘2-3_LK_control_interactions.omv’ on OSF: https://osf.io/e25p7/?view_only=c3b351f5764b42a7a626a1077537be88).

⁹In order to test whether the waitlist group improved more than the treatment group from T2-T3, the reference categories for the Condition and Time variables were set to ‘treatment group’ and ‘Time 2’ (rather than ‘waitlist group’ and ‘Time 1’). A significant Condition*Time interaction with a positive slope value means that the waitlist group improved significantly more than the treatment group from T2-T3. See file ‘2-1_LK_data_long_T2-T3d.omv’ on OSF: https://osf.io/e25p7/?view_only=c3b351f5764b42a7a626a1077537be88.

Table 3 Summary of mixed models results, regressing outcomes onto groups and time

	1a) Did TG increase T1-2?			1b) Did WG increase T1-2?			1c) Did TG increase more than WG T1-2?			2a) Did TG decrease T2-3?			2b) Did TG increase T1-3?			3a) Did WG increase T2-3?			3b) Did WG increase more than TG T2-3?		
	B	p	d	B	p	d	B	p	d	B	p	d	B	p	d	B	p	d	B	p	d
Kindness	0.53	<0.001	0.57	0.08	0.034	0.09	0.45	<0.001	0.49	-0.10	0.018	0.48	0.03	0.433	0.03	0.13	0.027	0.14			
Happiness	0.40	<0.001	0.42	0.06	0.167	0.06	0.34	<0.001	0.36	0.03	0.497	0.45	0.18	<0.001	0.19	0.15	0.015	0.16			
SDQ																					
Emotions	0.18	<0.001	0.33	-0.02	0.344	-0.04	0.20	<0.001	0.37	0.01	0.828	0.02	0.13	<0.001	0.24	0.13	<0.001	0.24			
Behaviour	0.16	<0.001	0.34	-0.01	0.524	-0.02	0.17	<0.001	0.36	-0.01	0.625	0.32	0.03	0.226	0.06	0.04	0.227	0.08			
Attention	0.27	<0.001	0.43	-0.03	0.222	-0.05	0.30	<0.001	0.48	-0.08	0.003	0.30	0.03	0.191	0.05	0.11	0.002	0.18			
Peers	0.18	<0.001	0.41	0.03	0.105	0.07	0.15	<0.001	0.34	0.03	0.131	0.48	-0.02	0.296	-0.05	-0.05	0.072	-0.12			
Prosociality	0.23	<0.001	0.41	0.02	0.487	0.04	0.21	<0.001	0.37	-0.04	0.075	0.32	0.03	0.260	0.05	0.07	0.041	0.13			
Academics																					
Maths	0.24	<0.001	0.24	0.04	0.359	0.04	0.20	<0.001	0.20	0.13	0.002	0.36	0.08	0.065	0.08	-0.05	0.424	-0.05			
Reading	0.24	<0.001	0.24	0.11	0.007	0.11	0.13	0.02	0.13	0.15	<0.001	0.39	0.14	<0.001	0.14	0.00	0.939	0.00			
Mean d	0.38		0.38		0.03		0.34		0.34		0.01	0.38		0.09							

TG: Treatment Group; WG: Waitlist Group; T1: Time 1; T2: Time 2; T3: Time 3. Bold; p ≤ 0.005

4 Discussion

The present study provides evidence for the positive effects of the Learn Kind curriculum on student outcomes. Compared to the waitlist control group, teachers in the treatment group reported that students were kinder—moving from an average kindness grade of C to B—and happier, and experienced fewer emotional problems, were better behaved, paid more attention in class, had better peer relations, showed more prosocial behaviour, and had better math and reading grades ($\bar{d} = 0.34$).

These findings are consistent with, and compare favourably to, the results of previous meta-analyses of SEL programs. For example, in this study, the improvement in emotions ($d = 0.37$) was greater than that found in previous research ($g = 0.14$, [4]; $g = 0.24$, [11]). The improvement in behaviour ($d = 0.36$) was greater than in previous research ($g = 0.16$ [4]). The improvement in peer relations ($d = 0.34$) was greater than in previous research ($g = 0.22$ [4]). The improvement in prosocial behaviour ($d = 0.37$) was at the upper end of the range of previous research ($g = 0.18$, [4]; $g = 0.23$, [11]; $d = 0.39$, [21]). The improvement in maths ($d = 0.20$) is lower than, but comparable to previous research ($ES = 0.26$, [5]). And the improvement in reading ($d = 0.13$) is lower than previous research ($ES = 0.25$, [5]). All told, despite its shorter duration, Learn Kind has effects that approach and in some cases exceed the threshold for impactful learning programs ($d = 0.40$) suggested by Hattie [16], and as such represents a promising choice for boosting SEL outcomes in schools.

We also found that these positive effects persisted for at least three months after the intervention ($\bar{d} = 0.38$). These follow-up effects compare favourably with the results of previous meta-analyses of SEL programs. For example, the improvement in emotions ($d = 0.33$) at 3+ months is comparable to the results of previous research on follow-up effects at 6+ months ($g = 0.12$, [4]) and 12+ months ($ES = 0.16$, [23]). The improvement in behaviour ($d = 0.32$) after three months is comparable to the results of previous research on follow-up effects at 6+ months ($g = 0.22$, [4]) and 12+ months ($ES = 0.14$, [23]). The improvement in peer relations ($d = 0.48$) after three months is comparable to the results of previous research on follow-up effects at 6+ months ($g = 0.27$, [4]). And the improvement in prosociality ($d = 0.32$) after three months is comparable to the results of previous research on follow-up effects at 6+ months ($g = 0.14$, [4], $d = 0.12$, [21]) and 12+ months ($ES = 0.13$, [23]). However, since there is no untreated control group at T3, these findings should be interpreted with care.

We also found that the results for the initial treatment group did not appear to be fully replicated in the waitlist group, which showed improvements only in emotions and reading. For the reasons stated above, in the absence of a genuine control group in the second part of the study (T2-3), it is not clear how these findings should be interpreted.¹⁰ Nevertheless, taking them at face value, we can still ask why there was this discrepancy. One possibility is that the effects of Learn Kind were genuinely different at different times of the school year. Perhaps teachers have to ‘catch students early’ and set the tone for the rest of the year, whereas attempting to teach kindness later in the year, when students

¹⁰Nor is it clear whether or how these results should be pooled with those of the initial treatment group. One possibility includes taking the average of the pre-post results (that is, averaging across the pre-post results for the treatment group T1-2, and the waitlist group T2-3). This gives an average effect size of $\bar{d} = 0.23$. Another is to take the average of the control and quasi-control results (that is, averaging across the difference between the treatment and waitlist groups T1-2, and the difference between groups T2-3). This gives an average effect size of $\bar{d} = 0.21$. These two ways of pooling the findings give similar results, that are still within the range of ‘small’ effects, and still compare favourably with previous research in most cases.

may be more fatigued, or when a pattern of relationships has already been established, is less successful. Further research will be needed to test this possibility.

Of course, we should keep in mind the limitations of the present study. The study was slightly underpowered; whereas we had aimed for 29 teachers per experimental group per time point, we achieved an average of 25 (and some groups consisted of 20). Although the teachers were randomly allocated to experimental conditions, they were nevertheless a self-selecting sample who elected to teach the curriculum, and elected to take part in the experiment. Although starting levels for most outcome variables are not significantly different between groups, the starting levels for three of the nine variables (kindness, behaviour, peer relations) were significantly different (at $p < 0.005$, controlling for multiple comparisons). Although other-reported data (by teachers) may have the advantage of being less subject to social desirability bias than self-reported data (by students) [3], it may have the disadvantage of being less accurate. The study did not investigate whether effects on behaviour at school generalized to behaviour outside of school (for example, in the home). The study did not investigate how the effects of Learn Kind interacted with the values and personality of the students (or their parents). The study did not investigate the specific effects of Learn Kind's teacher modules. And the results of the study pertain to one particular culture (the US), and so it is unclear whether the effects might be different in another culture—although we note that a recent review of SEL programs in 53 countries found no significant differences between “interventions with programs developed and delivered in the same country...compared to interventions implemented in a country different from where the program was developed” [4].

For these reasons, future research should also: use a larger sample; be conducted ‘school-wide’ with teachers randomly chosen to take part in the experiment; replicate the study with more fully randomized starting levels [20], collect student self- and parent other-reported data, of behaviour outside school; collect measures of student and parent values and personality; replicate the study with and without teacher modules; and replicate the study cross-culturally. It may also be informative to include a third ‘untreated’ control condition to provide a baseline set of measurements throughout.

Further research of this kind will enhance our understanding of kindness, and help to make the classroom a kinder place.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s44217-025-00689-y>.

Supplementary file1 (DOCX 2105 KB)

Acknowledgements

Thanks to our advisors and contributors who helped guide the design and development of the Learn Kind curriculum: Stephanie Valli, Keesa Addison, Lydia Phillips, Dipali Bhatt, Riley Bauling, and Kristen Robinson. And thanks to Ryan McManus for statistical advice.

Author contributions

MNT: conceptualization, methodology, formal analysis, data curation, writing—original draft, project administration CSM: formal analysis, data curation, writing—review & editing, visualization MP: conceptualization, software, resources RR: conceptualization, resources JL: conceptualization, resources, supervision, funding acquisition OSC: conceptualization, methodology, formal analysis, data curation, writing—review & editing, visualization, supervision.

Funding

This study was funded by Curtis and Shirley Chambers, The Hutchinson Family, Yukari Matsuzawa Pass, Greg Pass.

Data availability

Materials, data and analysis are available here: <https://osf.io/e25p7>.

Declarations

Ethics approval and consent to participate

Ethical and Independent Review Services (eandireview.com) determined that the project did not require IRB review using their Determination of Research with Human Subjects assessment process. The ethics determination is publicly available at https://osf.io/e25p7/?view_only=c3b351f5764b42a7a626a1077537be88. All methods were carried out in accordance with relevant guidelines and regulations. Participating teachers provided informed consent prior to survey completion.

Consent for publication

Participants were informed that anonymized data may be used in research and published.

Competing interests

All authors have been or currently are employed by the nonprofit organization, kindness.org. Compensation of kindness.org employees is not dependent on research output or findings. Financial donations were used to compensate participants and fund development of the Learnkind curriculum. Donors did not have any influence over the experimental design, data collection, research process, analysis, or writing of this manuscript.

Received: 10 December 2024 / Accepted: 7 July 2025

Published online: 27 November 2025

References

1. Baldwin CP, Baldwin AL. Children's judgments of kindness. *Child Dev.* 1970;41(1):29–47. <https://doi.org/10.2307/1127387>.
2. Blake PR, McAuliffe K, Corbit J, Callaghan TC, Barry O, Bowie A, Kleutsch L, Kramer KL, Ross E, Vongsachang H, Wrangham R, Warneken F. The ontogeny of fairness in seven societies. *Nature.* 2015;528(7581):258–61. <https://doi.org/10.1038/nature15703>.
3. Boon-Falleur M, Bouguen A, Charpentier A, Algan Y, Huillery É, Chevallier C. Simple questionnaires outperform behavioral tasks to measure socio-emotional skills in students. *Sci Rep.* 2022;12(1): Article 1. <https://doi.org/10.1038/s41598-021-04046-5>.
4. Cipriano C, Strambler MJ, Naples LH, Ha C, Kirk M, Wood M, Sehgal K, Zieher AK, Eveleigh A, McCarthy M, Funaro M, Ponnock A, Chow JC, Durlak J. The state of evidence for social and emotional learning: a contemporary meta-analysis of universal school-based SEL interventions. *Child Dev.* 2023;94(5):1181–204. <https://doi.org/10.1111/cdev.13968>.
5. Corcoran RP, Cheung ACK, Kim E, Xie C. Effective universal school-based social and emotional learning programs for improving academic achievement: a systematic review and meta-analysis of 50 years of research. *Educ Res Rev.* 2018;25:56–72. <https://doi.org/10.1016/j.edurev.2017.12.001>.
6. Cowell JM, Lee K, Malcolm-Smith S, Selcuk B, Zhou X, Decety J. The development of generosity and moral cognition across five cultures. *Dev Sci.* 2017. <https://doi.org/10.1111/desc.12403>.
7. Curry OS, Rowland LA, Van Lissa CJ, Zlotowitz S, McAlaney J, Whitehouse H. Happy to help? A systematic review and meta-analysis of the effects of performing acts of kindness on the well-being of the actor. *J Exp Soc Psychol.* 2018;76:320–9. <https://doi.org/10.1016/j.jesp.2018.02.014>.
8. Curry OS, San Miguel C, Tunç MN. The costs and benefits of kindness for kids. *J Exp Child Psychol.* 2024;246: 105987. <https://doi.org/10.1016/j.jecp.2024.105987>.
9. Curry OS, San Miguel C, Wilkinson J, Tunc M. Consistent Kindness: Money allocation and kind act decisions are regulated by a 'welfare trade-off ratio'. 2024. <https://osf.io/zuh59/>. [Accessed 6 September 2023]
10. Dawkins MB, Ting F, Stavans M, Baillargeon R. Early moral cognition: a principle-based approach. In: Poeppel D, Mangun GR, Gazzaniga MS, editors. *The cognitive neurosciences*. 6th ed. Cambridge: The MIT Press; 2020. p. 7–16. <https://doi.org/10.7551/mitpress/11442.003.0005>.
11. Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev.* 2011;82(1):405–32. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>.
12. Forster DE, Pedersen EJ, Smith A, McCullough ME, Lieberman D. Benefit valuation predicts gratitude. *Evol Hum Behav.* 2017;38(1):18–26. <https://doi.org/10.1016/j.evolhumbehav.2016.06.003>.
13. Goodman R. The strengths and difficulties questionnaire: a research note. *Child Psychol Psychiatry Allied Discip.* 1997;38(5):581–6. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>.
14. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry.* 2001;40(11):1337–45. <https://doi.org/10.1097/00004583-200111000-00015>.
15. Great Schools. K-12 school quality information and parenting resources. GreatSchools.Org. 2023. <https://www.greatschools.org/>. [Accessed 17 February 2023]
16. Hattie J. *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Routledge. 2008. <https://doi.org/10.4324/9780203887332>.
17. House BR, Silk JB, Henrich J, Barrett HC, Scelza BA, Boyette AH, Hewlett BS, McElreath R, Laurence S. Ontogeny of prosocial behavior across diverse societies. *Proc Natl Acad Sci.* 2013;110(36):14586–91. <https://doi.org/10.1073/pnas.1221217110>.
18. Ibbotson P. Little dictators: a developmental meta-analysis of prosocial behavior. *Curr Anthropol.* 2014;55(6):814–21. <https://doi.org/10.1086/679254>.
19. Lyubomirsky S, Sheldon KM, Schkade D. Pursuing happiness: the architecture of sustainable change. *Rev Gen Psychol.* 2005;9(2):111–31. <https://doi.org/10.1037/1089-2680.9.2.111>.
20. Owora AH, Dawson J, Gadbury G, Mestre LM, Pavela G, Mehta T, Vorland CJ, Xun P, Allison DB. Randomisation can do many things— but it cannot “fail.” *Significance.* 2022;19(1):20–3. <https://doi.org/10.1111/1740-9713.01609>.
21. Sklad M, Diekstra R, Ritter MD, Ben J, Gravesteyn C. Effectiveness of school-based universal social, emotional, and behavioral programs: do they enhance students' development in the area of skill, behavior, and adjustment? *Psychol Sch.* 2012;49(9):892–909. <https://doi.org/10.1002/pits.21641>.

22. Steger MF, Hicks BM, Kashdan TB, Krueger RF, Bouchard TJ. Genetic and environmental influences on the positive traits of the values in action classification, and biometric covariance with normal personality. *J Res Pers.* 2007;41(3):524–39. <https://doi.org/10.1016/j.jrp.2006.06.002>.
23. Taylor RD, Oberle E, Durlak JA, Weissberg RP. Promoting positive youth development through school-based social and emotional learning interventions: a meta-analysis of follow-up effects. *Child Dev.* 2017;88(4):1156–71. <https://doi.org/10.1111/cdev.12864>.
24. U.S. Census Bureau. INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS). [American Community Survey, ACS 5-Year Estimates Subject Tables]. 2022. <https://data.census.gov/table>. [Accessed 7 February 2024]
25. Wiggins GP, McGighe J. *Understanding by design*. ASCD; 2005.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.