



UNIVERSITY OF HELSINKI

<https://helda.helsinki.fi>

Development and initial validation of the THL Adverse Childhood Experiences Questionnaire (ACE-THL)

Hietamäki, Johanna; Laajasalo, Taina; Lindgren, Maija; Therman, Sebastian

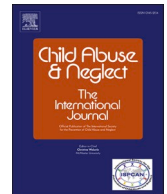
2023-12

Elsevier Ltd.

<http://hdl.handle.net/10138/569594>

Hietamäki, J, Laajasalo, T, Lindgren, M & Therman, S 2023, 'Development and initial validation of the THL Adverse Childhood Experiences Questionnaire (ACE-THL)', *Child Abuse and Neglect*, vol. 146, 106483. <https://doi.org/10.1016/j.chiabu.2023.106483>

Downloaded from Helda, University of Helsinki institutional repository. <https://helda.helsinki.fi>
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.



Development and initial validation of the THL Adverse Childhood Experiences Questionnaire (ACE-THL)

Johanna Hietamäki^{a,b,*}, Taina Laajasalo^{a,c}, Maija Lindgren^d, Sebastian Therman^d

^a Competence Cluster for Violence Prevention Work, Special Services Unit, Finnish Institute for Health and Welfare, Finland

^b Faculty of Social Sciences and Business Studies, University of Eastern Finland, Finland

^c Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

^d Mental Health, Public Health and Welfare, Finnish Institute for Health and Welfare, Finland

ARTICLE INFO

Keywords:

Adverse childhood experiences
Abuse
Neglect
Measurement
Instrument
Questionnaire

ABSTRACT

Background: The research on adverse childhood experiences (ACEs) has deepened our understanding of the long-lasting and cumulative effects of childhood adversities. However, the instruments measuring ACEs have several shortcomings, including limited item coverage, collapsing of items and response options, simplistic scoring, and inadequate psychometric assessments.

Objective: To design and conduct preliminarily psychometric testing for a brief new self-report instrument—the THL Adverse Childhood Experiences questionnaire (ACE-THL)—with a comprehensive set of clearly formulated items and appropriate response options.

Methods: A previously published process model was applied to develop the ACE-THL questionnaire, which was validated by cognitive interviews (N = 20). Interviewers and interviewees completed the questionnaire separately for a cross-informant comparison. In a separate survey panel validation, the respondents filled out the ACE-THL twice, two weeks apart (N = 513, with 426 in the follow-up). Interview data were used to improve item clarity, and test–retest reliability and structural validity were assessed with repeated survey data.

Results: The final 14-item questionnaire, including 12 ACE items and two items measuring protective experiences, was highly acceptable to the respondents. In the factor analysis of the quantitative data, a sufficiently single-dimensional construct was found, remaining stable in retesting two weeks later. The internal consistency (omega) of the a priori one-dimensional model was 0.89 and 0.90 at baseline and follow-up, respectively. The high test–retest reliability (mean score rank order correlation 0.93) of the ACE-THL indicated that the probed perceptions of childhood experiences are stable.

Conclusion: Based on the initial validation, the 14-item ACE-THL questionnaire is a reliable and valid instrument to measure adverse childhood experiences, as well as protective experiences.

1. Introduction

Since the seminal study of Felitti et al. (1998) and coining of the term adverse childhood experiences (ACEs), rapidly accumulating

* Corresponding author at: Competence Cluster for Violence Prevention Work, Special Services Unit, Finnish Institute for Health and Welfare, P.O. Box 30, 00271 Helsinki, Finland.

E-mail address: johanna.hietamaki@thl.fi (J. Hietamäki).

<https://doi.org/10.1016/j.chiabu.2023.106483>

Received 4 July 2023; Received in revised form 18 September 2023; Accepted 20 September 2023

Available online 1 November 2023

0145-2134/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

research evidence has deepened our understanding of the long-lasting and cumulative negative effects of childhood adversities. ACEs refer to a combination of violence, abuse, neglect, and household dysfunction, such as witnessing interparental violence, mental health and substance abuse problems in the family, imprisonment, or other types of separation from the parent. According to a large systematic review, reporting four or more of these experiences compared to reporting no ACEs doubles the risk of obesity and diabetes, triples the risk of smoking, cancer, heart, or respiratory disease, increases the risk of sexual risk behaviors and mental health and alcohol problems by about four times, and heightens the risk of drug abuse and violent behavior by as much as seven times (Hughes et al., 2017). In Europe, the annual ACE attributable costs were recently estimated to be equivalent to between 1 % (Sweden and Turkey) and 6 % (Ukraine) of the studied nations' gross domestic products (Hughes et al., 2021).

Studies have shown that the negative consequences for physical and mental health are not explained only by the health-damaging risk behaviors that follow ACEs, such as substance abuse or smoking, but also by other mechanisms that are still not fully understood (Linden & LeMoult, 2022). Several complementary theoretical explanations related to, for example, the effects of toxic stress and epigenetic modulation have been proposed (Miller et al., 2011; Shonkoff & Garner, 2012; Teicher et al., 2016). Because of the increased awareness of the consequences and costs of ACEs, several countries now systematically collect population-level data on the prevalence of ACEs to inform policymaking and intervention design with the aim of mitigating the consequences of trauma and chronic stress that may arise with exposure.

Childhood adversities and their consequences have been studied for decades, yet the study by Felitti et al. (1998) became known as the first ACEs study; their study was part of a larger epidemiological survey and measured seven categories of ACEs in its first wave (physical abuse, emotional abuse, sexual abuse, parental separation or divorce, witnessing violence against a mother, and household member mental illness or substance abuse and incarceration) and two additional ACEs in the second wave (physical neglect and emotional neglect). The items chosen to measure these 10 categories of ACEs were adopted from previous surveys and questionnaires (see Holden et al., 2020; Krinner et al., 2021). Since then, several studies have tried to improve the value of the original set of items by adding new adversities, such as financial difficulties, bullying, and other peer victimization while removing some of the original ones, which, in few cases, were based on psychometric testing utilizing different samples (e.g., Afifi et al., 2020; Finkelhor et al., 2013; Koita et al., 2018).

According to two reviews, at least 20 ACE instruments exist, varying in length (from under 10 to over 20 items), the number of categories assessed, the wording of the questions, and the types of response scales utilized (Bethell et al., 2017; Holden et al., 2020). Only four of the measures were designed for adult retrospective self-report (Bethell et al., 2017). Several shortcomings in the available set of ACE instruments have been identified, with the most common challenges relating to limited item coverage, collapsing of items and response options, simplistic scoring approach, and the lack of overall psychometric assessment, including reliability and validity testing (Holden et al., 2020; McLennan et al., 2020). Reidy et al. (2021) exemplified how altering measurement schemes impacts the interpretation of ACE-related research findings. When comparing the effect sizes for models measuring ACEs as binary versus ordinal, a significant variability in the effect sizes and predictive power of the individual ACEs was found. In addition, ACE instruments tend to focus solely on adversities, although it is known that protective factors and positive childhood experiences, such as experiences of benevolent interactions with trusted adults, buffer the effects of maltreatment or other adverse experiences (Bellis et al., 2017; Narayan et al., 2020). Joint inventories of ACEs and protective childhood experiences have been called for (Bethell et al., 2019). Based on the identified shortcomings, several researchers have acknowledged the need to urgently improve ACE instruments and the limited research base related to them (Holden et al., 2020; Reidy et al., 2021).

The aim of the present study was to design and conduct preliminary validation and testing for a short new questionnaire, with a comprehensive set of clearly formulated items and appropriate response options. The questionnaire is expected to have a sufficiently one-dimensional structure, high test-retest stability, and no more than minimal differential item functioning across age and gender groups. Using mixed methods, we cognitively interview a small sample to ensure face validity, sufficient scope, and acceptability of an

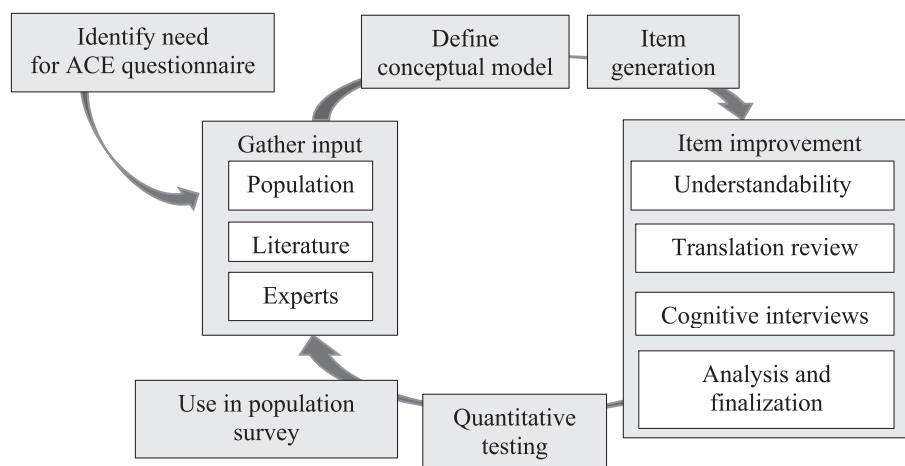


Fig. 1. Instrument development process (modified from Rothrock et al., 2011).

expert-generated selection of items, and psychometrically test a modified version in a larger sample of adults. The resulting Adverse Childhood Experiences–THL (ACE-THL), developed by the Finnish Institute for Health and Welfare (THL), is intended to be suitable for both the general adult population and for more targeted samples within the fields of psychiatry and social work.

2. Methods

We applied a modified version of Rothrock et al.'s (2011) model (Fig. 1), which has been designed to support the development of valid questionnaires with a mix of qualitative and quantitative methods. Following the recommended steps for item generation and improvement (Lee, 2006), the questionnaire validation consisted of two phases. First, in-person cognitive interviews were used to assess item clarity and how well the responses to the questionnaire covered the participant's experiences (individual interviews, $N = 20$), and second, reliability and structural validity were assessed with a repeated internet panel survey ($N = 510$).

2.1. Development of the questionnaire: item generation and improvement

To formulate the preliminary pool of essential items, we utilized a deductive method (e.g., Boateng et al., 2018) consisting of a literature review, assessment of existing scales, and their ACE indicators. Our conceptual model of ACEs was based on, but not restricted to, the 10 ACE categories used in several previous studies, meaning experiences such as peer rejection and financial struggles were considered relevant, in addition to abuse, neglect, and household dysfunction. After perusing the literature, six widely used international ACE measures were compared: the ACE Study questionnaire (Felitti et al., 1998), the CYW ACE-Q Teen questionnaire (Burke and Renschler, 2023), the Short ACE-IQ and SCM (Meinck et al., 2016), the BRFSS ACE (WHO, 2009), the ACE-IQ (WHO, 2018), and the TADS (Trauma and Distress Scale; Salokangas et al., 2016). The TADS was included because it has already been validated in Finland, addresses several ACEs, and uses an informative 5-point response scale.

Using an iterative expert consensus procedure, 150 items were reviewed. The reasons for excluding items were related to ambiguity and suboptimal formulation (e.g., double-barreled questions, redundancy, or nonessentiality). We included two positively framed items from the TADS related to emotional support. The inclusion of positive childhood experiences was justified based on the literature review. After formulating the first draft version of the questionnaire and pool of items, based on our own expert evaluation and observations derived from the literature, the exact formulation and wording of the instructions and items, as well as the type/scale of responses, were refined until a full consensus was reached among the reviewers. At the same time, we assessed the translatability of the questionnaire. Our group of five experts consisted of researchers (all with PhD degrees) in the fields of child abuse, domestic violence, forensic psychology, child protection, mental health, and psychometrics. In addition, we consulted a child psychiatrist with three decades of experience in child psychiatry and adverse childhood experiences.

The survey was conducted in Finnish and the ACE-THL questionnaire was translated into English by a qualified translator and then translated back into the Finnish by another qualified translator who was blinded to the original questionnaire. The researchers and a bilingual and bicultural external expert checked the translations after both stages. This back-translation process is recommended by most cross-cultural adaptation guidelines (Epstein et al., 2015).

2.2. Interview phase

The interviews were conducted by experienced researchers from a research company who invited interviewees from the participants enrolled in their national research panel.

Altogether, 20 voluntary respondents took part in the interview phase. Interviewees were selected from those who were registered in the research panel, considering the national representativeness in gender, age, socio-economic status, and family background in the recruitment process. The interview data consisted of 20 interviews, 20 self-report questionnaires, and 20 questionnaires filled out by the interviewers based on the interviews. The duration of each interview was approximately one hour. Half of the interviewees self-identified as women and half as men. The average age was 59 years (range 27–64 years). Most of the interviewees were employed (80 %); the rest were students (10 %) or retired (10 %). Household income varied from “10 000–20,000 €/year” to “over 81,000 €/year,” with the median income being “50,001–60,000 €/year.”

The ACE-THL questionnaire was first validated by cognitive interviews to elucidate the participants' thought processes as they responded to the questionnaire. In addition, we adopted a qualitative pretest interview method involving the interviewee as a co-expert to discover interpretations of the questionnaire and create suitable wordings (Buschle et al., 2022; Knafl et al., 2007; Willis, 2006).

The interviewee filled out the questionnaire independently before the interview. Next, the interviewer conducted a cognitive interview utilizing the thinking-aloud method and asked the respondent—one item at a time—what they thought when answering that item (Willis, 2006). By eliciting the thought processes that the interviewees associated with the items, it was possible to determine whether they had similar interpretations of the items as intended (Willis, 2006). Interviewees were asked not to say which answer option they had chosen. At the same time, the interviewer filled out the questionnaire based on the interview, which allowed for the consistency of the answers to be examined (Boateng et al., 2018). The interviewer also asked whether it was difficult to understand the text of the instructions, the items, or the response options and whether there were any suggestions for rewording questions to make them easier to understand (Buschle et al., 2022).

To systematically utilize the interview material, we adapted the guidelines of Willis (2006) and Knafl et al. (2007). The analysis focused on the changes from the original questionnaire and the reasons for these changes. The researchers went over the analysis

framework together to ensure that the practices were consistent. Each researcher (the authors) then analyzed five interviews by listening to the interview recordings and working with the fully transcribed interviews.

The transcripts were divided into quotes (text fragments) of one or a few sentences pertaining to a particular part of the instructions, an item, the response scale, or the questionnaire as a whole, and were coded as such. Each quote was also coded as being informed by the interviewee's personal experience or not. Notes were made of all proposal parts of the quotation and what kind of change was suggested. The researcher judged the need for change associated with each quote on an 11-point scale ($-5 = \text{extremely good/no changes needed}$, ..., $5 = \text{extreme need for change}$).

Quotes were then sorted by topic (question, response instruction, items, response scale, questionnaire) and need for change. The results were discussed together among the researchers, starting with the quotes associated with the greatest need for change, and consensus changes were made to the questionnaire.

Furthermore, the score distributions of the interviewees and their interviewers were compared, and inter-reporter correlations were calculated.

2.3. Quantitative survey

In the survey panel validation, the respondents were invited to fill out the ACE-THL twice online via the research company's website, approximately two weeks apart. A few additional questions were included regarding the perceived functionality and feasibility of responding to the questionnaire. Of the 2220 people invited, 513 (23 %) participated in the first round and, of them, 426 (83 %) participated in the second.

2.3.1. Panel participant self-selection and follow-up missingness

To identify any major self-selection bias in the results, we compared the demographic variable distributions of the participants and the 77 % of panel members who chose not to respond to the first round of the survey. We also compared the dropouts (17 %) to those who participated in the second round on demographics, as well as first-round responses. Because all these variables were ordinal and highly skewed, we used the nonparametric permutation Brunner–Munzel test of stochastic equality (Neubert & Brunner, 2007), as recommended by Karch (2021), and implemented in the *brunnermunzel* package version 2.0 (Ara, 2022) running in R 4.2.3 (R Core Team, 2023). For comparability across various variable types, we report the corresponding effect size, namely, the stochastic superiority index $\hat{p} = P(X_1 > X_2) + \frac{1}{2}P(X_1 = X_2)$, the probability that a random member of the first group has a value higher than that of a random member of the second group, which is equivalent to the area under the curve (AUC) in ROC analysis.

2.3.2. Structural validity

Childhood adversities are known to be interrelated, so the responses can be assumed to form a single latent factor (the a priori model) amenable to confirmatory factor analysis. However, items related to the same contexts (e.g., at home), themes (e.g., violence), and wordings are likely to have substantial residual correlations, requiring iterative exploratory adjustments by the greatest modification index. Therefore, the baseline dataset was split into two random halves for an exploratory discovery phase and validation phase. The models were also evaluated in the follow-up dataset.

Factor analyses were conducted in Mplus 8.8 (Muthén & Muthén, 2017) by treating variables as ordinal using the weighted least square mean and variance adjusted (WLSMV) estimators to estimate separate factor loadings and thresholds for all response options, with 100 random starts per estimation but otherwise standard settings. Model fit is reported with the CFI, RMSEA, and SRMR indices. Factor scores were computed with the *maximum* a posteriori method to allow for comparison with the scale mean scores. McDonald's omega (ω_{total}) was used as the index of the internal consistency of the selected factor model (McNeish, 2018).

2.3.3. Test–retest reliability

Test–retest reliability was assessed using the item-specific internal rank-order correlation comparing each respondent's responses across two rounds of responses. A correlation of 0.75 or higher was considered excellent, and a correlation of 0.6 or higher can be considered the minimum requirement for continued use of the items. Equivalent rank-order correlations were also calculated for the sum scores and factor scores.

In addition, responses were dichotomized into common (*often – almost always*) and less common (*never – sometimes*) experiences, with test–retest reliability assessed by the phi coefficient (r_ϕ); using this coarser measurement, a value of 0.4 or higher was considered a minimum requirement.

2.3.4. Differential item functioning

Because the questionnaire was intended for the same underlying construct of childhood adversity experience across various population groups, we examined differential item functioning (DIF, also known as measurement noninvariance) for individual statements across gender, age groups, household income, and region of residence. All multivalued group variables were dichotomized to obtain sufficient group sizes: age groups to approximate a median split, region into “capital region” or not, and household income to define a low-income group (household gross income up to 30,000 € per year), approximating the Eurostat definition of having below 60 % of the population median in disposable income. The comparisons used an iterative purification algorithm appropriate for ordinal items implemented in the *lordif* package version 0.3–3 (Choi et al., 2011) in R 4.2.3, employing a graded response model equivalent to the a priori model in the factor analyses. The package's default settings were used, and the group comparison DIF criterion was set to 2

% or more of item variance (as estimated with McFadden pseudo- R^2) being explained by the total DIF, which is the sum of uniform (constant across trait levels) and nonuniform (conditional on trait level) DIF. In each DIF analysis, the response categories of an item were collapsed as necessary to ensure a minimum of five cases per cell.

2.4. Ethical considerations

The research plan was approved by the Ethical Committee of the Finnish Institute for Health and Welfare. The participants were given detailed information about the study's purposes. The interviewees provided informed consent before they participated in the interviews. The informants were guaranteed anonymity throughout the study, and the questionnaire included information and contacts for easy-access services for emotional help. Each interviewee received a 40 € gift card as compensation for their time and effort. Both datasets were deidentified.

3. Results

3.1. Questionnaire modification

The questionnaire on adverse childhood experiences (ACE-THL) concerns childhood experiences before the age of 18. The questionnaire contains a total of 14 items. The original questionnaire contained 13 items, and this version was used in the interview validation. Based on the interviews, an item on sexual harassment was added to the questionnaire. The final ACE-THL questionnaire contains 14 statements (Appendix 1). Of the items, 12 concern ACEs: violence against the child (bullying, physical, emotional and sexual violence, violence between parents), neglect (physical neglect), and household dysfunction (mental illness, substance abuse, economic troubles, loss of a parent). Two items were related to protective factors such as supportive social connections. The ACE-THL uses a 5-point frequency response scale ranging from *Never* to *Constantly*.

Based on the analysis of the data from the interviews, we revised the introductory text, the scale, and 10 items, while three items (items 1, 10, and 14; Appendix 1) remained unchanged, and one item was added (item 3). The introductory text was shortened, and we added a note and validation of the sensitive nature of the topic (Appendix 1).

The overall questionnaire was highly acceptable to the respondents, and only a single person out of 513 reported discomfort in answering.

3.1.1. Introductory text, structure, and response scale

There were two challenges related to the introductory text and questionnaire. The interviewees reflected on the fact that 18 years may contain quite different periods of time in terms of experiences. It was considered whether the respondent should be guided to answer in general and average terms or, on the other hand, according to the worst episode, but it was decided that the respondents would get to decide by themselves how relevant they saw the experience. Another theme often discussed in the interviews was how life “really” was versus how one experienced it now as an adult; for example, as an adult, a person could interpret that their childhood was financially poor, though it did not feel like that as a child. No additional instruction related to the retrospective nature of the interpretation of the experiences was added.

The full range of the response scale was used in the data for nearly all items, indicating its suitability. Based on feedback, the wording of one response, *Almost always*, changed to *Constantly*. The same response scale was used for all statements. The structure of the questionnaire was divided into two sections because it was sometimes unclear to the interviewees which context each statement referred to (Appendix 1). The sections were *Anywhere, at home, or elsewhere* and *Parents and family*.

3.1.2. Items

The main contents of the items were understood in the ways intended by the researchers, but there was a need to clarify by modifying words and sentences. We also added one new item to assess sexual harassment or violence.

Two items did not sufficiently distinguish between the respondents; no one chose the extreme answer option, even though this would have been possible based on the interview data. The interviewees also pondered what kind of violence should be interpreted as “sufficiently serious.” Therefore, the term “serious violence” was replaced with “physical violence.” Many older respondents did not interpret disciplinary physical violence as violence in their childhood. Therefore, the words “pulled my hair” were added to the item to include disciplinary violence in the definition. All interviewees had experienced emotional violence at least rarely. Regarding shouting and swearing, they had difficulties distinguishing between isolated insensitive interactions and incidents of emotional violence, and these were deemed less severe than insults, belittling, and humiliation. The words “shouted or swore at me” were therefore removed from the item concerning emotional violence.

The item concerning sexual violence remained unchanged, but the interviewees raised the issue of noncontact sexual harassment and sexual harassment on social media, and these considerations led to the addition of one item: “I was subjected to sexual harassment or sexual violence.”

The term “confide in” was perceived as difficult to understand and outdated, and the term was replaced with “talk about my own things.”

In caregiver's intimate partner violence, the words "parents hurt each other" were replaced with "my other parent" because, in some cases, the original item led to the question of whether violence should be mutual. Furthermore, in the item on parental substance use, the word "intoxicated" was added based on interviewees' suggestions, and "drunk or on drugs" was moved inside parentheses. This clarification was added because interviewees used the word "intoxicated" and asked whether this refers to "moderate alcohol consumption" (e.g., a single glass of wine).

Words that limited the content of the item were rearranged or removed to make it more understandable. In the item concerning neglect, the word "school" was removed because it limited the response to certain days or a certain age (Item 11). Mental health problems in general were rearranged to ask first about mental health problems and later address specific examples of mental problems.

Overall, the interviewees felt that exploring difficult childhood experiences was meaningful. Those who had experienced ACEs found it easier to choose responses than those who had not experienced ACEs. Those who had no ACE experience wondered, for example, what the threshold for problematic substance use is.

There was a strong rank-order correlation between interviewer and interviewee responses to the same items, with correlation coefficients ranging from 0.62 to 0.91 (median 0.80). Absolute correlations between interviewer ratings and interviewee ratings to *other* items had a median of 0.22, showing that the high concordance was specific. Coefficients are shown in Supplementary Table 1.

3.2. Quantitative survey

3.2.1. Demographic information and follow-up missingness

The baseline respondent demographics are reported in Table 1 and response distributions in Table 2. Compared with the entire invited panel, the respondents were older ($\hat{p} = 0.68$, Brunner-Munzel statistic 13.88, $p < .001$), reported higher income brackets ($\hat{p} = 0.55$, B-M 3.64, $p < .001$), and were less often from the capital region ($\hat{p} = 0.54$, B-M 3.45, $p < .001$) and more often from the rest of southern Finland ($\hat{p} = 0.53$, B-M 2.79, $p = .005$) but had the same gender distribution. Compared with those who responded on the second round, follow-up dropouts were more often male ($\hat{p} = 0.58$; B-M 2.68, $p = .001$) and had more often experienced sexual harassment ($\hat{p} = 0.56$, B-M 2.43, $p = .016$; 71 % vs. 82 % *Never*) but did not differ in other baseline item response distributions, response time, use of mobile vs. desktop device, age, geographical distribution, or self-reported household income bracket. Though being somewhat skewed towards older adults, our survey sample was representative of the general population, making the parameters of our factor model and our differential functioning results generalizable.

3.2.2. Response distributions

The baseline response frequencies are shown in Table 2.

3.2.3. Structural validity

The fit of the a priori single-dimensional model was poor in the baseline discovery dataset, necessitating the freeing of four residual correlations to be estimated. This modified model performed well in both the validation and follow-up datasets, supporting the hypothesized unidimensionality of the scale, and all model reliabilities were excellent. Model fit and internal consistency indicators are reported in Table 3. The standardized factor model estimated from the combined discovery and validation datasets is shown in Table 4. The linear correlation between scale mean scores and factor scores was 0.96 both in the combined baseline dataset and in the follow-

Table 1
Baseline respondent demographics (N = 513).

Variable	Group	N	%
Gender	Female	291	56.7
Age	18–24	20	3.9
	25–34	130	25.3
	35–49	134	26.1
	50–64	148	28.8
	65–79	81	15.8
Household gross income, € per year	<10,000	35	6.8
	10,000–20,000	40	7.8
	20,001–30,000	55	10.7
	30,001–40,000	60	11.7
	40,001–50,000	47	9.2
	50,001–60,000	57	11.1
	60,001–70,000	34	6.6
	70,001–80,000	32	6.2
	80,001–90,000	29	5.7
	over 90,000	51	9.9
	(Income unavailable)	73	
Geographical region of residence	Capital (Helsinki & Uusimaa)	187	36.5
	South	120	23.4
	West	117	22.8
	North & East	89	17.3

Table 2
Baseline response frequencies.

Nr.	Item content	Item n	Never	Rarely	Sometimes	Often	Constantly
1	Bullied	511	10.6 %	32.9 %	36.2 %	16.0 %	4.3 %
2	Physical violence	507	45.6 %	31.4 %	19.9 %	2.8 %	0.4 %
3	Sexual harassment	510	72.9 %	13.3 %	11.2 %	2.2 %	0.4 %
4	Sexual touching	512	75.6 %	13.9 %	8.8 %	1.6 %	0.2 %
5	Adult confidant ^a	512	17.0 %	16.4 %	20.1 %	23.4 %	23.0 %
6	Parent humiliated me	512	51.0 %	22.5 %	18.6 %	6.4 %	1.6 %
7	Parent physically hurt me	512	41.4 %	32.2 %	22.5 %	3.7 %	0.2 %
8	Parent physically hurt other parent	512	78.1 %	8.8 %	9.4 %	3.7 %	–
9	Parent intoxicated	510	36.3 %	27.3 %	19.2 %	16.1 %	1.2 %
10	Financial trouble	511	33.1 %	24.9 %	25.8 %	11.0 %	5.3 %
11	Hunger/clothing	511	86.1 %	8.0 %	5.1 %	0.6 %	0.2 %
12	Family mental health problems	512	70.3 %	9.2 %	12.5 %	5.9 %	2.1 %
13	Loss of parent	512	75.4 %	5.5 %	11.1 %	6.4 %	1.6 %
14	Valued at home ^a	512	3.3 %	6.8 %	23.6 %	35.2 %	31.1 %

^a Positively worded item.

Table 3
Factor model fit and internal consistency.

Model	df	RMSEA	CFI	SRMR	Mean R ²	McDonald's ω_{total}
Discovery, one-dimensional	77	0.161	0.767	0.137	38.8 %	0.89
Discovery, adjusted	71	0.082	0.943	0.075	31.1 %	0.86
Validation, one-dimensional	67	0.150	0.802	0.109	38.6 %	0.89
Validation, adjusted	71	0.089	0.934	0.074	33.3 %	0.87
Follow-up, one-dimensional	66	0.153	0.819	0.112	39.7 %	0.90
Follow-up, adjusted	70	0.069	0.965	0.057	33.1 %	0.87

df = degrees of freedom; RMSEA = root mean square error of approximation.

CFI = comparative fit index; SRMR = standardized root mean square residual.

Table 4
Standardized parameters of the adjusted factor model in the combined baseline dataset.

Nr.	Item content	Loading	Thresholds			
			1	2	3	4
1	Bullied	0.44	−1.25	−0.17	0.83	1.72
2	Physical violence	0.47	−0.11	0.74	1.86	2.66
3	Sexual harassment	0.50	0.61	1.09	1.95	2.66
4	Sexual touching	0.47	0.69	1.25	2.11	2.89
5	Adult confidant ^a	−0.54	−0.95	−0.43	0.09	0.74
6	Parent humiliated me	0.84	0.02	0.63	1.41	2.15
7	Parent physically hurt me	0.55	−0.22	0.63	1.76	2.89
8	Parent physically hurt parent	0.67	0.78	1.12	1.79	–
9	Parent intoxicated	0.40	−0.35	0.35	0.94	2.27
10	Financial trouble	0.48	−0.44	0.20	0.99	1.62
11	Hunger/clothing	0.67	1.09	1.57	2.42	2.89
12	Family mental health problems	0.58	0.53	0.82	1.41	2.02
13	Loss of parent	0.38	0.69	0.87	1.41	2.15
14	Valued at home ^a	−0.72	−1.84	−1.27	−0.42	0.49

Residual correlations: items 1 & 2: 0.47; items 3 & 4: 0.86; items 5 & 14: 0.40; items 10 & 11: 0.68.

^a Positively worded item.

Table 5
Items with differential functioning.

Group comparison	Item	Content	Categories after collapsing	DIF R ²	
				Uniform	Nonuniform
Gender	2	Physical violence	4	0.04	0.01
	3	Sexual harassment	2	0.13	<0.001
	4	Sexual touching	2	0.11	<0.001
Age	12	Family mental health problems	3	0.05	<0.001

R² = McFadden pseudo-R² effect size.

up; the measurement accuracy of mean scores is thus comparable to that of factor scores, at least at non-extreme levels of ACEs.

3.2.4. Test-retest reliability

The rank-order correlation between the mean scores over the two time points was 0.93 and 0.92 between factor scores. The test-retest rank-order correlations for the individual items were also excellent, ranging from 0.84 (“Valued at Home”) to 0.98 (“Hunger/Clothing”), the median value being 0.91. With dichotomized responses, the r_ϕ coefficient was sufficient, ranging from 0.49 to 0.86 (median 0.72).

3.2.5. Differential item functioning

The DIF results are shown in Table 5. In the gender comparison, three items showed DIF above the threshold, namely “Physical Violence,” “Sexual Harassment,” and “Sexual Touching.” The physical violence item exhibited both uniform DIF, with males reporting more experiences compared with females than expected based on their overall level, and nonuniform DIF, with males over-represented on the less frequent end of the spectrum. In contrast, the sexual harassment and touching items evinced purely uniform DIF, with women reporting more of these experiences across all levels of overall ACEs.

Only the “Family Mental Health Problems” item exhibited age group DIF: those under 50 years of age reported higher frequencies for their ACE level (uniform DIF). There was no DIF by income or geographical region.

4. Discussion

The purpose of the present study was to describe the development of a new measure: the THL Adverse Childhood Experiences Questionnaire (ACE-THL). The development process, including psychometric testing, was designed to consider and avoid some of the pitfalls identified with previous instruments that had been used to measure ACEs.

The preliminary pool of items, formulated based on a literature review, an assessment of existing scales, and the ACE indicators of those scales, worked well. Based on comprehension, response processes and strategies, as well as ease of use, as assessed via semi-structured interviews, only a single item needed to be added, and small changes made to wordings. In the factor analysis of the quantitative data, a sufficiently single-dimensional construct was found, and was shown to be replicable in both a holdout and a follow-up sample. The high test-retest reliability of the ACE-THL indicated not only that the probed perceptions of childhood experiences were stable but that responses were nonrandom.

Though the one-dimensional model of the instrument was internally consistent, individual items varied greatly in their factor loadings, with “Loss of a Parent” and “Parent Intoxicated” having only acceptable shared variance with the whole, and “Parent Humiliated Me” and “Valued at Home” performing excellently. Item response distributions were also quite divergent, meaning that the use of a categorical factor model – rather than a sum or mean score – is warranted when estimating overall adversity, especially for measurement precision at range extremes.

Response propensity to some items differed between men and women, most evident in women having more frequent experiences of sexual harassment and unwanted sexual touching than men at the same overall ACE levels. This reflects the real effect of women being more often subjected to harassment (Kaltiala-Heino et al., 2018), which must be considered when forming summary indices and interpreting questionnaire results.

Respondents aged 50 or older indicated fewer familial mental health problems than younger respondents at equivalent ACE levels, perhaps reflecting a cohort effect of increased mental health literacy, improved diagnostics and treatment, and parent-child communication regarding this issue (Giano et al., 2020; Logan-Greene et al., 2014).

Even though Finland was the second country in the world to criminalize all types of violence against children in 1984 (Österman et al., 2014), in the cognitive interviews, the older respondents found it difficult to identify acts of disciplinary violence or corporal punishment as acts of violence. This was considered in the wording of the item related to physical violence at the less severe end of the violence continuum.

The ACE-THL covers the same domains as the original ACE questionnaire (physical, sexual, and emotional abuse, domestic violence and neglect, substance abuse or mental health problems in the family, and the loss of a parent), and additionally includes items such as bullying during childhood and financial difficulties, which were deemed to be important based on recent literature (e.g., Finkelhor et al., 2013). The questionnaire also included two items related to positive experiences (having an adult to speak to and the feeling of being valued at home) that are infrequently measured simultaneously with ACEs and that can also improve the response experience.

This instrument aims to respond to several criticisms raised in the literature. Importantly, it allows the assessment of the frequency of experiences (see, e.g., Lacey & Minnis, 2020) instead of dichotomous scoring, including questions on protective experiences that should be considered when assessing the consequences of ACEs (Guo et al., 2022). Still, there are measurement-related issues that go beyond the mere construction of instruments, and the instrument alone will not solve many of the problems that plague ACE research, such as neglecting the weighting of individual ACEs (Lacey & Minnis, 2020). These can be remedied by improving the types of analyses conducted on the data that the instrument provides. Furthermore, additional research is needed not only on the relative importance of various adversities but also on effective interventions that can mitigate the effects of these specific adversities (Finkelhor et al., 2015).

Methodologically sound approaches to appropriately measure ACEs are necessary to address ACEs as a public health issue. With a standard of measurement for examining ACEs, designing national prevention strategies and tailoring public health interventions becomes possible. This first version of the ACE-THL was developed to be used as part of a large forthcoming population-based survey related to adult health status and risks. Therefore, relevant demographic, social, and health behavior questions will be obtained from the survey (e.g., family structure, socioeconomic status, and various physical and mental health related variables). We will utilize the survey data to assess the different co-occurrences of ACEs and protective experiences, the weights that can be assigned to different experiences and their combinations regarding different outcome measures, and how chronicity of the experiences affects the outcomes. We will also be able to compare the usefulness of the latent factor and individual ACE items as predictors. Reidy et al. (2021) found that measuring ACEs individually, rather than as a sum score, improves the amount of variance explained in the overall outcome and that reliance on the total may hinder the attempts to identify which ACEs are the most salient to specific outcomes. Nonetheless, we acknowledge that there will be limitations to the survey data. Retrospective research cannot inform us about causality and recall bias must be considered. Baldwin et al.'s (2019) meta-analysis of 16 studies observed that the agreement between adult memories and childhood reports of abuse was poor, just over 50 %. A sizable proportion of childhood abuse is not recalled by adults, and on the other hand, many adults recall abuse that was not identified or reported when they were children. Thus, because prospective designs miss a proportion of cases, and unreported abuse may even be more severe (Kendall-Tackett & Becker-Blease, 2004), both retrospective and prospective study designs are needed. The focus group interviews, large sample size in the quantitative analysis, representative population sample, and comprehensive psychometric assessments are among the strengths of the present study. However, the present study has some limitations. The main limitation is the limited ethnic and cultural diversity of the study sample. It has been emphasized that ACE assessment has the possibility to reinforce social inequality if the contents and items are inappropriate for assessing adversity for minority groups (Mei et al., 2022).

4.1. Conclusion

Psychometric validation of ACE-related instruments is rare. We sought to implement the recommended questionnaire development methodology to ensure the validity of the ACE-THL, a novel self-report instrument for measuring adverse childhood experiences. Designed to be used in population-based surveys, ACE-THL comprises 14 items and adds to the existing outcome measures by measuring the frequency of the experiences, as well as including two items measuring protective, positive experiences. Psychometric properties, including factor model fit, measurement invariance, and item- and scale-level test-retest reliability, were excellent. Our future work will investigate whether some specific adverse experiences, either alone or in combination, are associated with adverse outcomes, and whether the frequency of experiences is related to the consequences. It is also important to examine the presence of protective factors and possible relationship between their frequency and the occurrence and consequence of adverse consequences. Finally, ACE-THL will be freely available for use. It is intended for research and surveillance purposes only, not for client screening, because the current evidence cautions against using ACE questionnaires as screening tools to predict individual risk (Baldwin et al., 2021; Finkelhor, 2018; Meehan et al., 2022).

Funding sources

-

Declaration of competing interest

None.

Data availability

The data that has been used is confidential.

Acknowledgements

Chief Specialist, Child Psychiatrist Jukka Mäkelä and Senior Specialist Johanna Cresswell-Smith

Appendix 1

ACE-THL web page: <https://thl.fi/lapsuuskokemuksia>. Copies of any publications using the ACE-THL should be sent to ace-thl@thl.fi to be added to the reference list for the method.

Childhood Experiences Questionnaire (ACE-THL)

Next, you will be asked about your childhood. You may find answering some of the questions distressing. The questions concern the time when you were **under 18 years old**.

Anywhere, at home or elsewhere:

	Never	Rarely	Sometimes	Often	All the time
1. I was bullied.	☐	☐	☐	☐	☐
2. I was a target of physical violence.	☐	☐	☐	☐	☐
3. I was a target of sexual harassment or sexual violence.	☐	☐	☐	☐	☐
4. Someone touched me in a sexual manner or made me touch them in a sexual manner, even though I didn't want to.	☐	☐	☐	☐	☐
5. I had an adult who I could talk to about my own matters.	☐	☐	☐	☐	☐

Parents and family:

A "parent" means an adult who was responsible for you . When responding, think about the parent who treated you the worst in relation to each statement.	Never	Rarely	Sometimes	Often	All the time
6. My parent insulted, belittled, or humiliated me.	☐	☐	☐	☐	☐
7. My parent slapped, hit, or kicked me, pulled my hair, or otherwise physically hurt me.	☐	☐	☐	☐	☐
8. My parent slapped, hit, kicked, pulled the hair of, or otherwise physically hurt my other parent.	☐	☐	☐	☐	☐
9. My parent was intoxicated (drunk or on drugs).	☐	☐	☐	☐	☐
10. My family had financial problems.	☐	☐	☐	☐	☐
11. I had to go hungry or, for example, wear dirty or ragged clothes.	☐	☐	☐	☐	☐
12. A member of my family had mental health problems, such as depression or self-destructiveness.	☐	☐	☐	☐	☐
13. I suffered from losing a parent, for example, because of separation, abandonment, or death.	☐	☐	☐	☐	☐
14. I was appreciated and considered important at home.	☐	☐	☐	☐	☐

© 2023, Finnish Institute for Health and Welfare (THL). This work is licensed under a CC BY 4.0 license.

Copies of any publications using the ACE-THL should be sent to ace-thl@thl.fi to be added to the reference list for the method.

The ACE-THL web page: <https://thl.fi/lapsuuskokemuksia>.

Appendix 2. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chiabu.2023.106483>.

References

- Afifi, T. O., Salmon, S., Garcés, I., Struck, S., Fortier, J., Taillieu, T., ... MacMillan, H. L. (2020). Confirmatory factor analysis of adverse childhood experiences (ACEs) among a community-based sample of parents and adolescents. *BMC Pediatrics*, 20(1), 178. <https://doi.org/10.1186/s12887-020-02063-3>
- Ara, T. (2022). brunnermunzel: (Permuted) Brunner-Munzel Test (2.0) [Computer software]. <https://CRAN.R-project.org/package=brunnermunzel>
- Baldwin, J. R., Caspi, A., Meehan, A. J., Ambler, A., Arseneault, L., Fisher, H. L., ... Danese, A. (2021). Population vs individual prediction of poor health from results of adverse childhood experiences screening. *JAMA Pediatrics*, 175(4), 385–393. <https://doi.org/10.1001/jamapediatrics.2020.5602>
- Baldwin, J. R., Reuben, A., Newbury, J. B., & Danese, A. (2019). Agreement between prospective and retrospective measures of childhood maltreatment: A systematic review and meta-analysis. *JAMA Psychiatry*, 76(6), 584–593. <https://doi.org/10.1001/jamapsychiatry.2019.0097>
- Bellis, M. A., Hardcastle, K., Ford, K., Hughes, K., Ashton, K., Quigg, Z., & Butler, N. (2017). Does continuous trusted adult support in childhood impart life-course resilience against adverse childhood experiences - a retrospective study on adult health-harming behaviours and mental well-being. *BMC Psychiatry*, 17(1), 110. <https://doi.org/10.1186/s12888-017-1260-z>

- Bethell, C., Jones, J., Gombojav, N., Linkenbach, J., & Sege, R. (2019). Positive childhood experiences and adult mental and relational health in a statewide sample: Associations across adverse childhood experiences levels. *JAMA Pediatrics*, 173(11), Article e193007. <https://doi.org/10.1001/jamapediatrics.2019.3007>
- Bethell, C. D., Carle, A., Hudziak, J., Gombojav, N., Powers, K., Wade, R., & Braveman, P. (2017). Methods to assess adverse childhood experiences of children and families: Toward approaches to promote child well-being in policy and practice. *Academic Pediatrics*, 17(7S), S51–S69. <https://doi.org/10.1016/j.acap.2017.04.161>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Burke, H. N., & Renschler, T. S. (2023, April 26). Center for Youth Wellness ACE-questionnaire (CYW ACE-Q child, teen, teen SR). [version 7/2015]. Center for Youth Wellness. <https://centerforyouthwellness.org/aceq-pdf/>
- Buschle, C., Reiter, H., & Bethmann, A. (2022). The qualitative pretest interview for questionnaire development: Outline of programme and practice. *Quality & Quantity*, 56, 823–842. <https://doi.org/10.1007/s11135-021-01156-0>
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8), 1–30. <https://doi.org/10.18637/jss.v039.i08>
- Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology*, 68(4), 435–441. <https://doi.org/10.1016/j.jclinepi.2014.11.021>
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., ... Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The adverse childhood experiences (ACE) study. *American Journal of Preventive Medicine*, 14(4), 245–258. [https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8)
- Finkelhor, D. (2018). Screening for adverse childhood experiences (ACEs): Cautions and suggestions. *Child Abuse & Neglect*, 85, 174–179. <https://doi.org/10.1016/j.chiabu.2017.07.016>
- Finkelhor, D., Shattuck, A., Turner, H., & Hamby, S. (2013). Improving the adverse childhood experiences study scale. *JAMA Pediatrics*, 167(1), 70–75. <https://doi.org/10.1001/jamapediatrics.2013.420>
- Finkelhor, D., Shattuck, A., Turner, H., & Hamby, S. (2015). A revised inventory of Adverse Childhood Experiences. *Child abuse & neglect*, 48, 13–21. <https://doi.org/10.1016/j.chiabu.2015.07.011>
- Giano, Z., Wheeler, D. L., & Hubach, R. D. (2020). The frequencies and disparities of adverse childhood experiences in the U.S. *BMC Public Health*, 20(1), 1327. <https://doi.org/10.1186/s12889-020-09411-z>
- Guo, S., O'Connor, M., Mensah, F., Olsson, C. A., Goldfeld, S., Lacey, R. E., ... Priest, N. (2022). Measuring positive childhood experiences: Testing the structural and predictive validity of the health outcomes from positive experiences (HOPE) framework. *Academic Pediatrics*, 22(6), 942–951. <https://doi.org/10.1016/j.acap.2021.11.003>
- Holden, G. W., Gower, T., & Chmielewski, M. (2020). Methodological considerations in ACEs research. In G. J. G. Asmundson, & T. O. Afifi (Eds.), *Adverse childhood experiences: Using evidence to advance research, practice, policy, and prevention* (pp. 161–182). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-816065-7.00009-4>
- Hughes, K., Bellis, M. A., Hardcastle, K. A., Sethi, D., Butchart, A., Mikton, C., ... Dunne, M. P. (2017). The effect of multiple adverse childhood experiences on health: A systematic review and meta-analysis. *The Lancet Public Health*, 2(8), e356–e366. [https://doi.org/10.1016/S2468-2667\(17\)30118-4](https://doi.org/10.1016/S2468-2667(17)30118-4)
- Hughes, K., Ford, K., Bellis, M. A., Glendinning, F., Harrison, E., & Passmore, J. (2021). Health and financial costs of adverse childhood experiences in 28 European countries: A systematic review and meta-analysis. *The Lancet. Public Health*, 6(11), e848–e857. [https://doi.org/10.1016/S2468-2667\(21\)00232-2](https://doi.org/10.1016/S2468-2667(21)00232-2)
- Kaltiala-Heino, R., Savioja, H., Fröjd, S., & Marttunen, M. (2018). Experiences of sexual harassment are associated with the sexual behavior of 14- to 18-year-old adolescents. *Child Abuse & Neglect*, 77, 46–57. <https://doi.org/10.1016/j.chiabu.2017.12.014>
- Karch, J. D. (2021). Psychologists should use Brunner-Munzel's instead of Mann-Whitney's U test as the default nonparametric procedure. *Advances in Methods and Practices in Psychological Science*, 4(2). <https://doi.org/10.1177/2515245921999602>
- Kendall-Tackett, K., & Becker-Blease, K. (2004). The importance of retrospective findings in child maltreatment research. *Child Abuse & Neglect*, 28(7), 723–727. <https://doi.org/10.1016/j.chiabu.2004.02.002>
- Knafl, K., Deatrick, J., Gallo, A., Holcombe, G., Bakitas, M., Dixon, J., & Grey, M. (2007). The analysis and interpretation of cognitive interviews for instrument development. *Research in Nursing & Health*, 30(2), 224–234. <https://doi.org/10.1002/nur.20195>
- Koita, K., Long, D., Hessler, D., Benson, M., Daley, K., Bucci, M., ... Burke Harris, N. (2018). Development and implementation of a pediatric adverse childhood experiences (ACEs) and other determinants of health questionnaire in the pediatric medical home: A pilot study. *PLoS One*, 13(12), Article e0208088. <https://doi.org/10.1371/journal.pone.0208088>
- Krinner, L. M., Warren-Findlow, J., Bowling, J., Issel, L. M., & Reeve, C. L. (2021). The dimensionality of adverse childhood experiences: A scoping review of ACE dimensions measurement. *Child Abuse & Neglect*, 121, 105270. <https://doi.org/10.1016/j.chiabu.2021.105270>
- Lacey, R. E., & Minnis, H. (2020). Practitioner review: Twenty years of research with adverse childhood experience scores - advantages, disadvantages and applications to practice. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 61(2), 116–130. <https://doi.org/10.1111/jcpp.13135>
- Lee, J. A. (2006). Constructing effective questionnaires. In J. A. Pershing (Ed.), *Handbook of human performance technology: Principles, practices, and potential* (3rd ed., pp. 760–779). Wiley. <https://doi.org/10.1002/pfi.20023>
- Linden, W., & LeMoult, J. (2022). Editorial perspective: Adverse childhood events causally contribute to mental illness – We must act now and intervene early. *Journal of Child Psychology and Psychiatry*, 63, 715–719. <https://doi.org/10.1111/jcpp.13541>
- Logan-Greene, P., Green, S., Nurius, P. S., & Longhi, D. (2014). Distinct contributions of adverse childhood experiences and resilience resources: a cohort analysis of adult physical and mental health. *Social Work in Health Care*, 53(8), 776–797. <https://doi.org/10.1080/00981389.2014.9442>
- McLennan, J. D., MacMillan, H. L., & Afifi, T. O. (2020). Questioning the use of adverse childhood experiences (ACEs) questionnaires. *Child Abuse & Neglect*, 101, 104331. <https://doi.org/10.1016/j.chiabu.2019.104331>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Meehan, A. J., Baldwin, J. R., Lewis, S. J., MacLeod, J. G., & Danese, A. (2022). Poor individual risk classification from adverse childhood experiences screening. *American Journal of Preventive Medicine*, 62(3), 427–432. <https://doi.org/10.1016/j.amepre.2021.08.008>
- Mei, X., Li, J., Li, Z. S., Huang, S., Li, L. L., Huang, Y. H., & Liu, J. (2022). Psychometric evaluation of an adverse childhood experiences (ACEs) measurement tool: An equitable assessment or reinforcing biases? *Health & Justice*, 10(1), 34. <https://doi.org/10.1186/s40352-022-00198-2>
- Meinck, F., Steinert, J. I., Sethi, D., Gilbert, R., Bellis, M. A., Mikton, C., ... Baban, A. (2016). *Measuring and monitoring national prevalence of child maltreatment: A practical handbook*. World Health Organization. <https://iris.who.int/handle/10665/343818>
- Miller, G. E., Chen, E., & Parker, K. J. (2011). Psychological stress in childhood and susceptibility to the chronic diseases of aging: Moving toward a model of behavioral and biological mechanisms. *Psychological Bulletin*, 137(6), 959–997. <https://doi.org/10.1037/a0024768>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén. <https://statmodel.com>
- Narayan, A. J., Atzl, V. M., Merrick, J. S., Harris, W. W., & Lieberman, A. F. (2020). Developmental origins of ghosts and angels in the nursery: Adverse and benevolent childhood experiences. *Adversity and Resilience Science*, 1, 121–134. <https://doi.org/10.1007/s42844-020-00008-4>
- Neubert, K., & Brunner, E. (2007). A studentized permutation test for the non-parametric Behrens-Fisher problem. *Computational Statistics & Data Analysis*, 51(10), 5192–5204. <https://doi.org/10.1016/j.csda.2006.05.024>
- Österman, K., Björkqvist, K., & Wahlbeck, K. (2014). Twenty-eight years after the complete ban on the physical punishment of children in Finland: Trends and psychosocial concomitants. *Aggressive Behavior*, 40(6), 568–581. <https://doi.org/10.1002/ab.21537>
- R Core Team. (2023). R: A language and environment for statistical computing (4.2.3) [Computer software]. R Foundation for Statistical Computing. <http://www.r-project.org>
- Reidy, D. E., Niolon, P. H., Estefan, L. F., Kearns, M. C., D'Inverno, A. S., Marker, C. D., & Merrick, M. T. (2021). Measurement of adverse childhood experiences: It matters. *American Journal of Preventive Medicine*, 61(6), 821–830. <https://doi.org/10.1016/j.amepre.2021.05.043>
- Rothrock, N. E., Kaiser, K. A., & Cella, D. (2011). Developing a valid patient-reported outcome measure. *Clinical Pharmacology and Therapeutics*, 90(5), 737–742. <https://doi.org/10.1038/clpt.2011.195>

- Salokangas, R. K. R., Schultze-Lutter, F., Patterson, P., von Reventlow, H. G., Heinimaa, M., From, T., ... Tuominen, L. (2016). Psychometric properties of the trauma and distress scale, TADS, in an adult community sample in Finland. *European Journal of Psychotraumatology*, 7, 30062. <https://doi.org/10.3402/ejpt.v7.30062>
- Shonkoff, J. P., Garner, A. S., & Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, & Section on Developmental and Behavioral Pediatrics. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, 129(1), e232–e246. <https://doi.org/10.1542/peds.2011-2663>
- Teicher, M. H., Samson, J. A., Anderson, C. M., & Ohashi, K. (2016). The effects of childhood maltreatment on brain structure, function and connectivity. *Nature Reviews Neuroscience*, 17(10), 652–666. <https://doi.org/10.1038/nrn.2016.111>
- WHO. (2009). *Addressing adverse childhood experiences to improve public health: Expert consultation, 4–5 May 2009*. World Health Organization. http://www.aceresponse.org/img/uploads/file/who_aces_2009.pdf
- WHO. (2018). *Adverse childhood experiences international questionnaire (ACE-IQ)*. World Health Organization. [https://www.who.int/publications/m/item/adverse-childhood-experiences-international-questionnaire-\(ace-iq\)](https://www.who.int/publications/m/item/adverse-childhood-experiences-international-questionnaire-(ace-iq))
- Willis, G. (2006). Cognitive interviewing as a tool for improving the informed consent process. *Journal of Empirical Research on Human Research Ethics*, 1(1), 9–24. <https://doi.org/10.1525/jer.2006.1.1.9>