The many faces of early life adversity - Content overlap in validated assessment instruments as well as in fear and reward learning research

Alina Koppold¹, Julia Ruge¹, Tobias Hecker² ³, and Tina B. Lonsdorf¹ ⁴

¹University Medical Center Hamburg-Eppendorf, Department of Systems Neuroscience, Hamburg, Germany
²University of Bielefeld, Clinical Psychology & Violence Research, Bielefeld, Germany
³University of Bielefeld, Institute of Interdisciplinary Conflict and Violence Research, Bielefeld, Germany
⁴University of Bielefeld, Biological Psychology and Cognitive Neuroscience, Bielefeld, Germany

The precise assessment of childhood adversity is crucial for understanding the impact of aversive events on mental and physical development. However, the plethora of assessment tools currently used in the literature with unknown overlap in childhood adversity types covered hamper comparability and cumulative knowledge generation. In this study, we conducted two separate item-level content analyses of in total 35 questionnaires aiming to assess childhood adversity. These include 13 questionnaires that were recently recommended based on strong psychometric properties as well as additional 25 questionnaires that were identified through a systematic literature search. The latter provides important insights into the actual use of childhood adversity questionnaires in a specific, exemplary research field (i.e., the association between childhood adversity and threat and reward learning). Of note, only 3 of the recommended questionnaires were employed in this research field. Both item-wise content analysis illustrate substantial heterogeneity in the adversity types assessed across these questionnaires and hence highlight limited overlap in content (i.e., adversity types) covered by different questionnaires. Furthermore, we observed considerable differences in structural properties across all included questionnaires such as the number of items, age ranges assessed as well as the specific response formats (e.g., binary vs. continuous assessments, self vs. caregiver). We discuss implications for the interpretation, comparability and integration of the results from the existing literature and derive specific recommendations for future research. In sum, the substantial heterogeneity in the assessment and operationalization of childhood adversity emphasizes the urgent need for theoretical and methodological solutions to promote comparability, replicability of childhood adversity assessment and foster cumulative knowledge generation in research on the association of childhood adversity and physical as well as psychological health.

Keywords: measurement heterogeneity, content analysis, early childhood adversity, childhood maltreatment, questionnaires

Introduction

Exposure to childhood adversity has been established as a strong risk factor and predictor for the development of psychopathology during the last decades (Anda et al., 2006; Danese & Widom, 2021; Felitti et al., 1998; Gilbert et al., 2009; C. Heim & Nemeroff, 2001; C. M. Heim et al., 2019; McLaughlin et al., 2015; Teicher et al., 2021). In particular, exposure to childhood adversity is rather common - approximately 60% of all children and adolescents are exposed to at least one adverse event (Madigan et al., 2023) - it is linked to substantial individual suffering and societal costs (Hughes et al., 2021). Childhood adversity has been defined in the literature as experiences that necessitate substantial adaptation by the child and deviate from the expected environment based on the assumption that appropriate input is fundamental for the normative (neuro-)developmental maturation (Cicchetti & Lynch, 1995; McLaughlin et al., 2021). In the literature, childhood adversity is assessed through a variety of approaches including official records, interviews or questionnaires (as either self-report or caregiver/parent report versions). Research on childhood adversity and in particular cumulative knowledge generation in the field is, however, hampered by measurement and operationalizational challenges (Elson et al., 2023; Flake & Fried, 2019; Flake et al., 2017). As for general trauma, there
is a lack of an agreed-up definition as well as substantial heterogeneity in assessment instruments. More precisely, it has recently been demonstrated that a multitude of different measures exist for the assessment of general trauma that differ in the number and type of specific trauma types assessed (Karstoft & Armour, 2022). As a consequence, individuals classified as exposed in one study, may not be classified as exposed in another study. This assessment heterogeneity may in part originate from definitions evolving over time. More precisely, the criteria to define “trauma”, a necessary precondition for the diagnosis of post-traumatic stress disorder (PTSD), differ between editions of diagnostic criteria (e.g., DSM-III, DSM-IV, ICD-11, Frueh et al., 2004; Karstoft & Armour, 2022; Slep et al., 2015). As a consequence, different definitions of (presumably) the same construct (i.e traumatic experiences), are at the base of different questionnaires used for the assessment of general trauma experience. Similarly, standardized definitions and assessment instruments of childhood adversity are lacking (Slep et al., 2015). For instance, the DSM-5 and DSM-IV place childhood adversity categories (e.g., physical or sexual abuse) under V codes (i.e., non-mental disorder conditions). The ICD-10 includes child maltreatment exposures in the Z codes (similar to DSM’s V codes), T codes (injury), and Y codes (external causes of injury and death). However, none of these codes (T, V, Y, Z) are operationalized by clear criteria for defining exposure (Slep et al., 2015), cover predominantly physical events while generally neglecting socio-emotional adverse events and do not discriminate between exposure and experience (McLaughlin et al., 2021). Relatedly, childhood adversity questionnaires developed early often focused on physical and sexual violence (e.g., Castelda et al., 2007; Felitti et al., 1998; Lynch & Cicchetti, 1998), while more recently developed assessment tools also cover additional facets such as emotional violence, neglect, peer, and/or witnessed violence (Teicher & Parigger, 2015). Both meta-scientific research and clinically-oriented research highlighted that such heterogeneity in measurement and assessment tools is a key threat for comparability, clinical translation and cumulative science alike (Elson et al., 2023; Flake & Fried, 2019; Flake et al., 2017). For instance, prospective and retrospective assessment instruments of childhood adversity did not identify the same individuals as “exposed” (Baldwin et al., 2019). Yet, the extent to which differences in content (i.e., adversity types) or differences in structural properties (e.g., response format, age ranges, valence of the experience) between different questionnaires contribute to heterogeneity in the literature remains unclear to date. To this end, a systematic investigation and illustration of measurement heterogeneity is an important first step to “ensure that we understand what we talk about when we talk about “childhood adversity” (cf. Karstoft & Armour, 2022) and hence putting this important research field on a solid foundation. In light of the co-existence of a plethora of distinct questionnaires for the assessment of general childhood adversity, it is a challenge for researchers, particularly those new to the field, to select the optimal assessment tool for their purposes from this rich buffet. Hence, the overarching aim of this work is to provide guidance by establishing a detailed overview on the currently used measures. To this end, we aim to inform decision-making, facilitate comparisons, and contribute to the cumulative generation and integration of knowledge from various findings.

More precisely, this study comprises three specific aims. The first aim involves examining the extent of item-based content overlap (i.e., childhood adversity types covered) between childhood adversity questionnaires that have been recommended based on strong psychometric properties. The second aim is to provide insight into item-based content overlap between childhood adversity questionnaires used in an exemplary research field (i.e., associations between childhood adversity and experimental threat and reward learning (Ruge et al., 2023). Finally, the third aim is to provide a comparative and comprehensive description and overview of the structural characteristics of the identified childhood adversity assessment questionnaires. Together, these results will allow us to distill clear considerations and recommendations for future research and provide much needed guidance in interpreting and integrating the current literature.

Method

1. Selection of Questionnaires

Only self-report and caregiver-report measures focusing on exposure to adverse events during childhood and adolescence were included. Official reports, interviews, and measures focusing on stress rather than adverse events were not considered in line with the approach of previously published content analyses (e.g., Karstoft and Armour, 2022 on General Trauma; Fried, 2017 on depression). A detailed list of all questionnaires is provided in Supplementary Table 1.

a. General Content Analysis.

First, we aimed to investigate the item-based overlap between questionnaires, which i) capture adverse childhood exposures, ii) are the most recurrently validated questionnaires from 2010 to 2020 (as reported in Georgieva et al., 2022) and iii) show strong psychometric properties in a recent systematic review (inter-
nal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, criterion validity, or responsiveness, Saini et al., 2019). We included all questionnaires that met at least 3 strong to moderate criteria, as determined by the Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) checklist in both publications (Georgieva et al., 2022; Saini et al., 2019). In the study by Saini et al. (2019), a total of 10 questionnaires (CTF-long form, CTQ-short form, MACE, CATS; ETI-SR, AE-III, CCMI, BCAS, ETI, CEVQ) fulfilled these criteria. In the study by Georgieva et al. (2022), 5 questionnaires (CTQ-short form, MACE, CAPI, IPARAN, PAT2.0) met the specified criteria. The overlap of two questionnaires led to the inclusion of a total of 13 questionnaires. Both studies (Georgieva et al., 2022; Saini et al., 2019) relied on the childhood adversity definition of the WHO Consultation on Child Abuse Prevention stating: ‘Child abuse or maltreatment constitutes all forms of physical and/or emotional ill-treatment, sexual abuse or neglect, treatment or commercial or other exploitation, resulting in actual or potential harm to the child’s health, survival, development or dignity in the context of a relationship of responsibility, trust or power (World Health Organization, 2024)’, which we hence also adopt here. Ten questionnaires were openly available online in the form of articles or dissertations; an additional two were obtained directly from the original authors, and one more was downloaded from a specific platform (ePROVIDE, https://eprovide.mapitrust.org).

b. Field Specific Content Analysis.

Second, we aimed to investigate the item-based overlap between questionnaires which capture childhood adversity in “the field” using this specific research field as a case example. As learning is a central mechanism through which environmental inputs shape emotional and cognitive processes and ultimately behavior, learning mechanisms are key candidates potentially underlying the biological embedding of exposure to childhood adversity and its impact on development and risk for psychopathology (McLaughlin & Sheridan, 2016). Thus, we chose research on the association between childhood adversity and threat as well as reward learning (Ruge et al., 2023) as a case example. We extracted the questionnaires used to assess childhood adversity from the studies identified through a recent systematic review (Ruge et al., 2023). In brief, 25 different questionnaires were extracted from a total of 73 studies (threat: N=36, reward: N=37). Fifteen of the identified 25 questionnaires were included in the Jaccard overlap analyses that provide a metric about the item-based content overlap between pairs of measures (details in the section ‘Statistical Analyses’ below). We did not consider assessment instruments that focus on clinical diagnoses, low family socioeconomic status based on income or postcode and parental substance abuse only, questionnaires focussing on specific types of exposure only, such as threatening experiences only [“Child Abuse Potential Inventory” (CAPI, Milner, 1994), the Violence Exposure Scale Revised child version (VEX-R, Fox & Leavitt, 1995), the Violence Exposure Scale Revised parent version (VEX-R_parent, Fox & Leavitt, 1995), Exposure to Violence (ETV, Selnér-O'Hagan et al., 1998), Conflict Tactics Scale (CTS, Straus et al., 1998)], deprivation only [Home Screening Questionnaire (HSQ, Frankenburg & Coons, 1986), Multidimensional Neglectful Behavior Scale (MNBS, Kantor et al., 2004), MacArthur SSS Scale of Subjective Social Status (MacArthur SSS Scale, Adler et al., 2000)] or bullying in school [Generalized Harassment Questionnaire (GHQ, Radoman et al., 2019), Peer Experiences Scale (PEQ, Casement et al., 2014)]. Additionally, the Traumatic Events Inventory (TEI) was excluded, as in the studies that employed the TEI, only a subset of three out of a total of 15 items was used (Morrison et al., 2022; Rowland et al., 2022). We did not include assessment instruments from which only a subset of questions was used in a specific study while no information was provided on which questions were included (Casement et al., 2014). Sixteen of the included questionnaires were openly available online in form of an article or dissertation, five were obtained from the original authors, and one was downloaded from specific platform (ePROVIDE, https://eprovide.mapitrust.org). Some questionnaires have undergone revisions or updates, e.g., the Life events checklist (LEC, Gray et al., 2004) has been modified to align with DSM-5 standards, resulting in the LEC-5. For our analyses, we used the exact version that was employed in the study of the example field.

2. Content Analyses

First, we extracted all items across questionnaires which resulted in 833 items for the general content analysis and 759 for the field-specific content analysis. Second, an item-based content overlap analysis (Fried, 2017; Karstoft & Armour, 2022) was conducted. In brief, based on the conservative approach used by Fried (2017) and Karstoft and Armour (2022), we identified matching adversity types across questionnaires. Two raters (AK, JR) independently determined the type of adversity of each item. First, based on two questionnaires, the two coders independently generated a pool of exposure types, guided by the subscales of the questionnaires. Subsequently, we compared our exposure type pools and reached an agreement. The established
pool of exposure types was then applied to all subsequent questionnaire items. Afterwards the coders were unblinded with respect to the results of the second coder and resolved mismatching assignments of items to exposure types collaboratively. Difficult decisions were discussed with the team members and decisions were made collaboratively. More precisely, if items were relatively similar and refer to the same adversity type, such as the item “There was someone to make sure you had a safe place to stay” (HTQ) and the item “Did someone force you to do things you didn’t want to do?” (PC-CTS), they were coded as the same adversity type (i.e., physical neglect). Furthermore, categories were labeled as specific as possible to differentiate between different albeit related events. For instance, some questionnaires assess specific adversity types using a single broadly formulated item (e.g., sexual abuse in the LEC), while other questionnaires provide a more fine-grained assessment through multiple items referring to the same adversity type but different perpetrators (e.g., sexual abuse by household person or sexual abuse by a peer in the MACE) or adversity subtypes (e.g., sexual abuse non-specific and sexual harassment in the MACE). As suggested by Pollingstad et al. (2005), “emotional abuse” was operationalized here as including verbal abuse, and hence verbal abuse is not listed as a separate adversity category. Items regarding a child’s behavior (e.g., sleep difficulties) as well as items about parental educational attitudes (e.g., “ignoring your child can be beneficial at times”) were categorized as ‘not event related’ (see Figure 5,6) as these cannot be directly linked to potential maltreatment. A significant number of items did not cover a single specific content but were formulated more broadly covering a broader range of contents (e.g. events). In these cases, a single item was scored more than once (for details see Figure Description 5 and 6). For instance, the item “Seeing someone die suddenly or get badly hurt or killed” (THS) was scored twice, once as content for ‘witnessed murder’ and once for ‘witnessed violence’.

3. Statistical Analyses

The agreement in the classification of item-based content between both raters was assessed using Cohen’s Kappa (unweighted, Watson and Petrie, 2010) in R Studio, utilizing the ‘irr’ package (Gamer et al., 2019). Jaccard-index based Item-wise Content Analysis.

The procedure as described in the following was identical for both the general content analysis and the field-specific analysis. In brief, the Jaccard Index, a widely used similarity coefficient for binary data, was calculated to estimate content overlap, which can range from 0 (no overlap) to 1 (complete overlap). The Jaccard Index, also known as the Jaccard similarity coefficient, is computed using the formula,

\[
s = \frac{s}{u_1 + u_2 + s}
\]

where s denotes the number of shared items between two questionnaires, and \(u_1\) and \(u_2\) represent the items that are exclusive to questionnaire1 and questionnaire2, respectively. As there is no widely accepted criterion for defining a weak or strong Jaccard similarity coefficient, the correlation coefficient rule proposed by Fried (2017) based on Evans (1996) is used here as benchmarks: a coefficient of 0.00-0.19 is considered very weak, 0.20-0.39, weak, 0.40-0.59 moderate, 0.60-0.79 strong, and 0.80-1.0 is considered very strong. The questionnaires used in the field-specific analysis are based on original studies from the systematic review (Ruge et al., 2023). In these original studies, the questionnaires were sometimes modified by the authors (e.g., by adding or removing items). In these cases (CLES, LEC, MNBS-CR, HSQ, GHQ, THQ, ETV, TEI, see Supplementary Table 1), the number of items included in the field-specific Jaccard overlap analysis corresponds to the number of items used in the respective studies, not the number of items in the original questionnaire. For the general content analysis, four questionnaires were excluded from Jaccard analyses due to their exclusive focus on a particular type of childhood adversity: the “Psychosocial Assessment Tool” (PAT2.0, Pai et al., 2008), tailored specifically for parents of children diagnosed with cancer; the “Identification of Parents at Risk for Child Abuse and Neglect” (IPARAN, Bouwmeester-Landweer, 2016), designed to assess parenting stress risk in parents with newborns; the “Child Abuse Potential Inventory” (CAPI, Milner, 1994); and the “Childhood Experiences of Violence Questionnaire” (CEVQ, Walsh et al., 2008), both designed to evaluate threat-related childhood adversity.

4. Comparative Overview: Descriptive Structural Characteristics of Childhood Adversity Questionnaires

To provide an overview of structural and format differences between questionnaires, we included all questionnaires from both content analyses, resulting in a total of 35, with 3 questionnaires overlapping (Figure 1). For each questionnaire, we extracted detailed information on the versions available and used in the content analyses described above. This included details on the response format, target population, the age range for which the original questionnaire was developed, valence and frequency of exposure, calculation of sum scores (quantification of exposure based
on responses), and available cut-off recommendations. Regarding questionnaires from the field-specific section, we recorded information from both the original questionnaire and any deviations found in the respective studies (e.g., number of items). For data analyses and visualizations as well as for the creation of the manuscript, we used the following R packages\(^1\). Data and code are available on Zenodo (doi: https://doi.org/10.5281/zenodo.10695485).

**Results**

**Inter-rater Agreement on Item-based Content**

High agreement between rater 1 and rater 2 with respect to the item-based content were observed for both the general content analysis (Cohen’s Kappa unweighted, \(\kappa = 0.81\)) and the field-specific content analysis (\(\kappa = 0.86\)).

**Overview of Descriptive Structural Characteristics**

The descriptives of the structural properties of the questionnaires revealed major differences between the instruments. First, the questionnaires differed significantly in their number of items. The smallest number of items was observed for the MacArthur SSS Scale, a deprivation-specific scale with 2 items, while the Assessing Environments had a maximum of 170 items (see Figure 2A). Second, the modality of the questionnaires varied from self-report, to caregiver report or self and caregiver report combined, semi-structured interview to age-specific self-report for adolescents, with self-reports generally being the most common modality (see Figure 2B). Notably, self-reports included both child self-reports prospectively and adult self-reports retrospectively (Figure 4B). Third, the purpose and target group varied from screening tools to diagnostic measures for adversity, such as age, frequency and duration of exposure is captured in at least one characteristic by 26 out of all 35 questionnaires. More precisely, the age of exposure is assessed by 17 questionnaires, whereas 2 questionnaires focus specifically on the last 12 months (PC-CTQ and GHQ, for details on corresponding questionnaires see Supplementary Table 1). The duration or frequency is assessed by 18 questionnaires, and the experienced valence of the event by 8 questionnaires, moreover 2 questionnaires, the LEC and UCLA PTSD-R, capture the role of the respondent in the event (victim, witnessed, learned about, Figure 3C).

**General Content Analysis**

Thirteen questionnaires assessing a total of 40 different adversity types were identified. Individual questionnaires captured between 7 (17.5%, CAPI) and 22 (55%, CCMI) of these adversity types (see Figure 5). The average item-based content overlap aiming to assess general childhood adversity (as calculated by the Jaccard Index for 9 out of the 13 questionnaires) ranged from 19% to 36%, indicating rather low similarity of adversity types across questionnaires (see Supplementary Table 2 for details). Overlap across questionnaires revealed an overall mean Jaccard Index of .29. At a descriptive

\(^{1}\text{R (Version 4.2.2; Team, 2022) and the R-packages ade4 (Bougeard and Dray, 2018; Chessel et al., 2004; Dray and Dufour, 2007, Version 1.7.20; Dray et al., 2007), bookdown (Version 0.30; Xie, 2016), corrplot2021 (Wei & Simko, 2021), data.table (Version 1.14.6; Dowle and Srinivasan, 2022), dplyr (Version 1.1.2; Wickham et al., 2023), effectsize (Version 0.8.2; Ben-Shachar et al., 2020), flexibletable (Version 0.9.4; Goel and Skintzos, 2023), forcats (Version 0.5.2; Wickham, 2022a, ggalluvial-article (Brunson, 2020), ggplot2 (Version 3.4.0; Wickham, 2016), ggsankey (Version 0.0.99999; Sjoberg, 2023), gridGraphics (Version 0.5.1; Murrell and Wen, 2020), gt (Version 0.8.0; Iannone et al., 2022), irr (Version 0.84.1; Gamer et al., 2019), kableExtra (Version 1.3.4; Zhu, 2021), knitr (Version 1.41; Xie, 2015), IpSolve (Version 5.6.17; Berkelaar et al., 2022), officer (Version 0.6.3; Goel, 2023), patchwork (Version 1.1.2; Pedersen, 2022), purrr (Version 1.0.1; Wickham and Henry, 2023), rcartocolor (Version 2.0.0; Nowosad, 2018), RColorBrewer (Version 1.3.1; Neuwirth, 2022), readr (Version 2.1.3; Wickham et al., 2022), readxl (Version 1.4.1; Wickham and Bryan, 2022), stringr (Version 1.5.1; Wickham, 2022b), tibble (Version 3.2.1; Müller and Wickham, 2023), tidyverse (Version 1.3.2; Wickham and Girlich, 2022), tidyverse (Version 1.3.2; Wickham et al., 2019), tinylabels (Version 0.2.3; Barth, 2022), viridis (Version 0.6.2; Garnier et al., 2021, 2022), and viridisLite (Version 0.4.1; Garnier et al., 2022Garnier et al., 2022)
Figure 1

The illustration highlights which questionnaires (for details see Supplementary Table 1) were recommended based on strong psychometric properties in the literature (general content analysis, Georgieva et al., 2022; Saini et al., 2019) and observed in the field in studies included in a systematic literature search on the association between childhood adversity and threat as well as reward learning processes (field specific content analysis) or both (overlap, Figure 1). A color within the color scheme represents a specific questionnaire, each of which is indicated above the circle by its abbreviation in a box.
Figure 2

Histogram A illustrates the distribution of the number of items in each questionnaire. Individual questionnaires differed markedly in the number of items included (2 - 170 items, Figure 2A), inherently resulting in lower Jaccard Indices. Some studies in the field-specific content analysis, only used a subset of items of a respective questionnaire (for details see Supplementary Table 1), however, the number of items depicted here reflects the original number of items, whereas the field specific content analysis is based on the items used in the respective study in “the field”. In total, 8 questionnaires (CLES, LEC, MNBS-CR, HSQ, GHQ, THQ, ETV and the TEI) were used in a modified version in the studies included in this review and hence included in the content analysis in this modified version. More specifically, for the CLES only 40 items out of originally 53 were used, for the modified version of the LEC (Scharfenort et al., 2016), items 1-3, 16, 17,19, 21, 24 were not used, for the MNBS-CR (Machlin et al., 2019; Milojevich et al., 2020), items 41-51 were not used, for the HSQ, items 6, 22, 29, 30, 34 were not used, for the GHQ, a modified version of the Generalized Workplace Harassment Questionnaire with 29 items, only 21 were used, for the THQ, only the last 6 items were used, for the ETV, only 13 out of 25 items were used, and for the TEI, only items 11, 12, 13 were used (Morrison et al., 2022; Rowland et al., 2022). For generation of the content analysis, we employed the modified versions of the questionnaire as reported in the publications included here. The extent of heterogeneity in assessment modalities, such as self-report and caregiver self-report (as shown in Figure 2B), was found to be quite diverse across questionnaires.

level, the BCAS had the lowest overlap with other questionnaires (.19) and CCM had the largest overlap with all other questionnaires (.36; see Supplementary Table 2). Pairwise comparisons between the questionnaires, (Table 3), revealed - at a descriptive level, the lowest degree of overlap was between the CATS and the CTQ-SF (.14) as well as BCAS and CCM, and the highest degree of overlap between the CTQ-LF and the AE-III (.58).

Field-specific Content Analysis

Twenty-two questionnaires assessing a total of 50 different adversity types were identified. Individual questionnaires captured between 1 (2%, MacArthur SSS Scale, PEQ) and 25 (50%, LSC-R) of these adversity types (see Figure 6). The average item-based content overlap aiming to assess general childhood adversity (as calculated by the Jaccard Index for 15 out of the 25 questionnaires) ranged from 1% to 34%, indicating rather low similarity of adversity types across questionnaires (see Supplementary Table 3 for details). Overlap across questionnaires revealed an overall mean Jaccard Index of .25. At a descriptive level, the PEQ had the lowest overlap with other questionnaires (.01), followed by the CECA-Q (.15). The TAQ had the highest overlap with all other questionnaires (.34; see Supplementary Table 3). At a descriptive level, pairwise comparisons between the questionnaires revealed that the lowest degree of overlap was between the PEQ and all other questionnaires (0.00), except the ACE (0.05), while the highest degree of overlap was observed between the UCLA-PTSD and the THQ (0.67). Interestingly, also in the general content analysis the CTQ (short form) revealed the lowest overlap with other questionnaires. Moreover, a study focusing on general adversity questionnaires (Karstoft & Armour, 2022) also identified the
Sankeys illustrates the frequency of different age ranges based on the original questionnaire. Most refer to 0-17 years, beyond that a large part of the questionnaires can be applied spanning lifetime (A). The score ranges of the questionnaires are usually obtained by adding up responses of the items (in some questionnaires these are additionally assigned to subcategories and weighted, e.g., MACE). Notably, no single score range occurs more than once. Two values in a row indicate that for this questionnaire exists more than one scoring procedure (B). Specific characteristics of exposure to adversity (C) revealed that for 54.3 % (19/35) questionnaires the age of exposure was assessed, 45.7 % (16/35) age of assessment, 22.8 % (8/35) the valence of exposure, and 51.4 % (18/35) the frequency of exposure. For the HSQ that asks about the child's environment in depth (e.g., 'How often does someone get a chance to read stories to your child?') where these specific exposure questions were not relevant, they were coded as 'no'. Notably, in two questionnaires the age of exposure refers only to specific time points (JVT, ETV) or the previous year (PC-CTQ, GHQ). For detailed corresponding questionnaire information see Supplementary Table 1.
Figure 4

Sankey illustrates responses variations from verbal options, i.e., dichotomous, open field questions, to different Likert scale versions whereas some capture frequency some not and/ or specify the time frame of exposure (A). Note that option refers to a predefined set of answer choices. The majority of questionnaires were retrospective self-reports for adults (B). Questionnaires aiming to prospectively assess childhood varied substantially in modality (self vs. caregiver report) and purpose (e.g., at risk, maltreatment screening). Note: pt LS = point Likert Scale.

THQ as showing the highest overlap (.59) in content overlap analyses.
Table 1

Jaccard similarity index matrix across questionnaires. Values range from 0 (no overlap) to 1 (total overlap). Mean Jaccard index across questionnaires was .29, indicating weak overlap. Minimum mean overlap was .19 (BCAS) and maximum mean overlap .36 (CCMI; for details see Supplementary Table 2).

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>CATS</th>
<th>CTQ-SF</th>
<th>ETI</th>
<th>ETISR-SF</th>
<th>MACE</th>
<th>BCAS</th>
<th>CCMI</th>
<th>CTQ-LF</th>
<th>AE-III</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATS</td>
<td>0.14</td>
<td>0.50</td>
<td>0.36</td>
<td>0.17</td>
<td>0.18</td>
<td>0.32</td>
<td>0.20</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>CTQ-SF</td>
<td>0.14</td>
<td>0.27</td>
<td>0.32</td>
<td>0.26</td>
<td>0.17</td>
<td>0.33</td>
<td>0.41</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>ETI</td>
<td>0.50</td>
<td>0.27</td>
<td>0.57</td>
<td>0.23</td>
<td>0.21</td>
<td>0.38</td>
<td>0.22</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>ETISR-SF</td>
<td>0.36</td>
<td>0.32</td>
<td>0.57</td>
<td>0.21</td>
<td>0.24</td>
<td>0.32</td>
<td>0.29</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>MACE</td>
<td>0.17</td>
<td>0.26</td>
<td>0.23</td>
<td>0.21</td>
<td>0.15</td>
<td>0.46</td>
<td>0.38</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>BCAS</td>
<td>0.18</td>
<td>0.17</td>
<td>0.21</td>
<td>0.24</td>
<td>0.15</td>
<td>0.14</td>
<td>0.19</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>CCMI</td>
<td>0.32</td>
<td>0.33</td>
<td>0.38</td>
<td>0.32</td>
<td>0.46</td>
<td>0.14</td>
<td>0.48</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>CTQ-LF</td>
<td>0.20</td>
<td>0.41</td>
<td>0.22</td>
<td>0.29</td>
<td>0.38</td>
<td>0.19</td>
<td>0.48</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>AE-III</td>
<td>0.09</td>
<td>0.30</td>
<td>0.15</td>
<td>0.25</td>
<td>0.44</td>
<td>0.24</td>
<td>0.45</td>
<td>0.58</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5

Co-occurrence of the 40 different childhood adversity types identified in the 13 questionnaires (with a total of 833 items) as recommended in (Georgieva et al., 2021; Saini et al., 2019). Three (7.5%) adversity types were only included in one individual questionnaire. Numbers refer to the event types listed below the circle. Colored dots in this figure represent the questionnaires listed in the legend on the right. Fully colored dots for an event type indicate that the respective questionnaire included more than a single item for this event type. Dots that are not filled indicate that only a single item captures this event in the respective questionnaire. Items highlighted in blue font indicate that this item was not filled out by the child but a caretaker (e.g., parent). In case a questionnaire item referred to two or more possible events (e.g., “witnessing severe illness OR death of a close person”) and the exposure to of one of these events was sufficient to indicate the experience of childhood adversity, the item was coded as present twice in order to represent all possible content in the analysis. This was the case for the CCMI 7 items and AE-III 3 items.
Table 2

Jaccard similarity index matrix across questionnaires. Values range from 0 (no overlap) to 1 (total overlap). Mean Jaccard index across questionnaires was 0.25, indicating weak overlap. Minimum mean overlap was 0.01 for PEQ and maximum mean 0.34 for TAQ (for details see Supplementary Table 3).

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>CECA.Q</th>
<th>CTQ-SF</th>
<th>ETISF-SR</th>
<th>JVQ</th>
<th>LEC</th>
<th>LSC</th>
<th>Life events’ checklist</th>
<th>THS</th>
<th>UCLA-PTSD</th>
<th>ACE</th>
<th>CLES</th>
<th>ELSQ</th>
<th>TAQ</th>
<th>THQ</th>
<th>PEQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CECA.Q</td>
<td>0.50</td>
<td>0.28</td>
<td>0.16</td>
<td>0.05</td>
<td>0.18</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
<td>0.17</td>
<td>0.10</td>
<td>0.14</td>
<td>0.25</td>
<td>0.09</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>CTQ-SF</td>
<td>0.50</td>
<td>0.32</td>
<td>0.24</td>
<td>0.10</td>
<td>0.25</td>
<td>0.04</td>
<td>0.10</td>
<td>0.09</td>
<td>0.26</td>
<td>0.14</td>
<td>0.18</td>
<td>0.33</td>
<td>0.13</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>ETISF-SR</td>
<td>0.28</td>
<td>0.32</td>
<td>0.20</td>
<td>0.27</td>
<td>0.38</td>
<td>0.21</td>
<td>0.42</td>
<td>0.38</td>
<td>0.36</td>
<td>0.30</td>
<td>0.35</td>
<td>0.61</td>
<td>0.41</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>JVQ</td>
<td>0.16</td>
<td>0.24</td>
<td>0.20</td>
<td>0.21</td>
<td>0.31</td>
<td>0.09</td>
<td>0.27</td>
<td>0.35</td>
<td>0.38</td>
<td>0.12</td>
<td>0.32</td>
<td>0.26</td>
<td>0.37</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>LEC</td>
<td>0.05</td>
<td>0.10</td>
<td>0.27</td>
<td>0.21</td>
<td>0.27</td>
<td>0.12</td>
<td>0.47</td>
<td>0.42</td>
<td>0.10</td>
<td>0.17</td>
<td>0.26</td>
<td>0.25</td>
<td>0.53</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>LSC</td>
<td>0.18</td>
<td>0.25</td>
<td>0.38</td>
<td>0.31</td>
<td>0.27</td>
<td>0.39</td>
<td>0.28</td>
<td>0.34</td>
<td>0.37</td>
<td>0.29</td>
<td>0.37</td>
<td>0.47</td>
<td>0.37</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Life events’ checklist</td>
<td>0.05</td>
<td>0.04</td>
<td>0.21</td>
<td>0.09</td>
<td>0.12</td>
<td>0.39</td>
<td>0.24</td>
<td>0.22</td>
<td>0.22</td>
<td>0.32</td>
<td>0.30</td>
<td>0.20</td>
<td>0.25</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>THS</td>
<td>0.05</td>
<td>0.10</td>
<td>0.42</td>
<td>0.27</td>
<td>0.47</td>
<td>0.28</td>
<td>0.24</td>
<td>0.53</td>
<td>0.15</td>
<td>0.29</td>
<td>0.33</td>
<td>0.36</td>
<td>0.56</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>UCLA-PTSD</td>
<td>0.10</td>
<td>0.09</td>
<td>0.38</td>
<td>0.35</td>
<td>0.42</td>
<td>0.34</td>
<td>0.22</td>
<td>0.53</td>
<td>0.22</td>
<td>0.36</td>
<td>0.38</td>
<td>0.67</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>ACE</td>
<td>0.17</td>
<td>0.26</td>
<td>0.36</td>
<td>0.38</td>
<td>0.10</td>
<td>0.37</td>
<td>0.22</td>
<td>0.15</td>
<td>0.22</td>
<td>0.31</td>
<td>0.46</td>
<td>0.46</td>
<td>0.17</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>CLES</td>
<td>0.10</td>
<td>0.14</td>
<td>0.30</td>
<td>0.12</td>
<td>0.17</td>
<td>0.29</td>
<td>0.32</td>
<td>0.29</td>
<td>0.26</td>
<td>0.31</td>
<td>0.35</td>
<td>0.42</td>
<td>0.24</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>ELSQ</td>
<td>0.14</td>
<td>0.18</td>
<td>0.35</td>
<td>0.32</td>
<td>0.26</td>
<td>0.37</td>
<td>0.30</td>
<td>0.33</td>
<td>0.36</td>
<td>0.46</td>
<td>0.35</td>
<td>0.46</td>
<td>0.33</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>TAQ</td>
<td>0.25</td>
<td>0.33</td>
<td>0.61</td>
<td>0.26</td>
<td>0.25</td>
<td>0.47</td>
<td>0.20</td>
<td>0.36</td>
<td>0.38</td>
<td>0.46</td>
<td>0.42</td>
<td>0.46</td>
<td>0.36</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>THQ</td>
<td>0.09</td>
<td>0.13</td>
<td>0.41</td>
<td>0.37</td>
<td>0.53</td>
<td>0.37</td>
<td>0.25</td>
<td>0.56</td>
<td>0.67</td>
<td>0.17</td>
<td>0.24</td>
<td>0.33</td>
<td>0.36</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>PEQ</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
**Exploratory Analyses**

Based on a reviewer’s comment, we also explored the content overlap within caregiver and self-report questionnaires by conducting a further content analysis. For caregiver questionnaires, a mean Jaccard Index of 0.22 across these questionnaires (CAPI, PC-CTS, VEX-R Caregiver Version, HSQ, CLES, IPARAN, PAT) was observed. Supplementary Table 4 lists the indices specific to each pair of caregiver questionnaire comparisons. We refer to Supplementary Figure 2 for an illustration of adversity types included in questionnaires separated for questionnaires using a) retrospective self report b) self report in pediatric samples as well as c) caregiver report in pediatric samples. The Jaccard indices for self-report questionnaires are illustrated in Table 2.

**Discussion**

Heterogeneity in measurement as well as assessment tools has been highlighted as a key threat for comparability between studies and cumulative science alike (Elson et al., 2023; Flake & Fried, 2019; Flake et al., 2017). Here we investigated item-based content overlap (i.e., adversity types) between different questionnaires used to assess childhood adversity. This will provide a first step for establishing a solid and common ground for future research on this important predictive risk factor for the development of both mental and physical disorders. More precisely, we focus on item-based content overlap between i) validated questionnaires that have been recommended based on strong psychometric properties (n=13, Georgieva et al., 2021; Saini et al., 2019) as well as on item-based content overlap between questionnaires used in one translationally highly relevant exemplary research field (e.g., associations between childhood adversity and threat as well as reward learning, n=24 questionnaires). It is noteworthy that only 3 of the recommended questionnaires, namely CAPI, ETI-SR, and CTQ-SF, were in fact employed in the exemplary research field of associations between threat and reward learning and childhood adversity. The ETI-SR and CTQ-SF capture a broad range of different adversity types during childhood (CTQ-SF: 10 types, ETI-SR: 15 types; see Figure 6) through retrospective self-reports in adults while the CAPI is an established questionnaire for the assessment of parents’ tendency towards physical abuse of children. These specific three questionnaires may be in fact used by scientists in practice, because they are fast to complete (according to the authors <15 min), provide clear scoring procedures for the evaluation, are available in multiple languages, and are freely available - at least the CTQ-SF and ETI-SR (MacDonald et al., 2016; Thombs et al., 2007). Limited Item-based Content Overlap across Childhood Adversity Questionnaires Both the general and field-specific item-based content analyses illustrate a limited overlap and hence substantial heterogeneity in content (i.e., adversity types) and number of adversities between the different questionnaires. None of the pairwise comparisons between two questionnaires yielded a content overlap of 0.60 or higher (mean content overlap .29 and .25 in the general and field-specific item-wise content analyses, respectively). This rather weak overlap among questionnaires indicates that different childhood adversity questionnaires should be compared with caution and may not be used interchangeably although this is often done in practice (see Ruge et al., 2023). In fact, different questionnaires may not even measure the same underlying “experience”. These results from childhood adversity questionnaires are in line with recent reports of low content overlap in assessment instruments of general trauma (Karstoft & Armour, 2022) as well as previous reports on limited agreement in the classification of “exposed” and “unexposed” individuals across different childhood adversity questionnaires - for instance between the CTQ and the KERF (german version of the MACE questionnaire, kappa .39, Koppold et al., 2023).

**Substantial Heterogeneity in Childhood Adversity Types across Questionnaires**

Our results across both content analyses also provide a valuable overview on the different adversity types considered as childhood adversity across questionnaires (i.e., general content analysis: n=40, field-specific content analysis: n=50, see Figure 5 and 6), which provides valuable conceptual information for the field. In addition, our results allow us to extract how frequently individual childhood adversity types (i.e., item-based content) are included across questionnaires. More precisely, the the most consistently included content across questionnaires were sexual abuse non-specific' (20 questionnaires) and 'physical abuse non-specific' (17 questionnaires), while 'sexual abuse peer' (2 questionnaires), 'financial problems' (in household, 4 questionnaires), or 'family crime' (6 questionnaires) were only included in a minority of questionnaires. In providing a comprehensive overview on the item-based content of childhood adversity questionnaires, we aim to facilitate the identification of questionnaires particularly suitable for specific research questions by making decision-relevant information accessible to the community (e.g., specificity and breadth of items, not covered content). It is particularly striking that childhood adversity types listed as the most common worldwide (United Nations Population Fund, 2024), such as child...
Co-occurrence of the 50 different childhood adversity types identified in the 24 questionnaires (with a total of 750 items) derived from the publications included in the systematic literature search (Ruge et al., 2023). Six (12%) adversity types were only included in one individual questionnaire. Numbers refer to the event types listed below the circle. Colored dots in this figure represent the questionnaires listed in the legend on the right. Fully colored dots for an event type indicate that the respective questionnaire included more than a single item for this event type. Dots that are not filled indicate that only a single item captures this event in the respective questionnaire. In case a study did not use all items of a questionnaire (see Supplementary Table 1), these are not included here. Items highlighted in blue font indicate that this item was not filled out by the child but a caretaker (e.g., parent). In case a questionnaire item referred to two or more possible events (e.g., “witnessing severe illness OR death of a close person”) and the exposure to one of these events was sufficient to indicate the experience of childhood adversity, the item was coded as present twice in order to represent all possible content in the analysis. This was the case for the ETISR-SF 3 items, LSC-R 2 items, Life events checklist 1 item, CLES 1 item, ACE-Q 1 item, HSQ 2 items, TAQ 1 item, THQ 1 item, UCLA 3 items, THS 1 item. In total, 8 questionnaires (CLES, LEC, MNBS-CR, HSQ, GHQ, THQ, ETV and the TEI) were used in a modified version in the studies included in this review and hence included in the content analysis in this modified version. More specifically, for the CLES only 40 items out of originally 53 (Smith & Pollak, 2021) were used, for the modified version of the LEC (Scharfenort et al., 2016), items 1-3, 16, 17, 19, 21, 24 were not used, for the MNBS-CR (Machlin et al., 2019; Milojevich et al., 2020), items 41-51 were not used, for the HSQ (Milojevich et al., 2020), items 22, 29, 30, 34 were not used, for the GHQ, a modified version of the Generalized Workplace Harassment Questionnaire (Radoman et al., 2019) with 29 items, only 21 were used, for the THQ (Young et al., 2019), only the last 6 items were used, for the ETV (Estrada et al., 2020), only 13 out of 25 items were used, and for the TEI, only items 11, 12, 13 were used (Morrison et al., 2022; Rowland et al., 2022). For generation of this figure as well as the content analysis, we employed the modified versions of the questionnaire as reported in the publications included here.
Figure 7

The barplot (A) illustrates, in a simulated sample of 50 subjects, each of whom experienced one of the 50 identified exposure types from the field-specific analysis (Figure 6), how many of the 50 individuals would be identified by each questionnaire. The “most diverse” exposure list would therefore be the LSC-R, which would only identify half of the exposed subjects. Exposure-specific questionnaires such as the MacArthur SSS Scale or the PEQ naturally perform even worse. Since some questionnaires weight responses to create sum-scores that determine whether certain cutoffs are exceeded, in B we depict the problem of forming sum scores for (1) the identification of exposed individuals and furthermore the fallacy of deriving severity from sum-scores.

marriage and child prostitution are not covered in any of 35 questionnaires included here. This impressively but also depressingly highlights a cultural gap and a potentially systematic bias. Furthermore, institutionalization, estimated to impact approximately 3.18 million to 9.42 million children worldwide (Desmond et al., 2020), is addressed in merely 6 out of the 35 included questionnaires, covered with only a single item each. This is particularly problematic as deprivation experiences due to institutionalization are highlighted prominently as a key adversity type in a central theoretical framework on the consequences of childhood adversity (i.e., DMAP, Sheridan and McLaughlin, 2014). In fact, our results show that in field specific research, deprivation is often assessed as single event types (such as official records of adoption or institutionalization, low family SES) without further characterization of individuals' exposure to or experience of other childhood adversity types (Ruge et al., 2023). This is particularly problematic because this approach lacks specificity to conclude derivation-specific effects, even though it's often done, and in addition hampers comparability of studies. In addition, it can be questioned if conceptualizing institutionalization or adoption as an adverse event and more specifically a deprivation experience is a valid approach. For instance, it can be questioned whether a person that has been adopted (i.e., exposure) has ever in fact experienced deprivation experiences. Similarly, conceptualizing the exposure to institutionalization as deprivation experience neglects the increased risk of institutionalized children to be exposed to threat-related or other adverse experiences and hence the exposure to institutionalization may not be linked specifically and exclusively to deviation experiences (see below 'fuzzy categories,' Smith and Pollak, 2022).

Potential Reasons for limited Item-based Content Overlap

Several considerations on potential reasons for the overserved generally low item-based content overlap between different questionnaires used to assess childhood adversity need to be discussed. First, despite all aiming to assess childhood adversity generally, some of the questionnaires included were - in part - designed for different purposes (e.g., screening, prospective risk assessment, retrospective report), different target groups (e.g., caregiver, exposed individual, pediatric sample vs. adult sample), or different target contexts (e.g., high
In addition to heterogeneous content covered across childhood adversity questionnaires, differences in potentially relevant structural characteristics may impact on the interpretation and results in the literature. More precisely, a range of response formats was employed including verbal report, open field questions, Likert-type, pictures, and binary (yes/no; true/false) response types (for details see Supplementary Table 1). Different response formats are well known to impact the answer for a number of reasons and biases (for details see Menold and Bogner, 2014). In addition to response formats, it is noteworthy that only one quarter of the included questionnaires provided scoring or cut-off recommendations (see Supplementary Table 1). Furthermore, these procedures vary even for a single questionnaire, for instance with respect to scoring recommendations (Life Stressor Checklist-Revised, Wolfe et al., 2012) or cut-off criteria (Childhood Trauma Questionnaire, Bernstein et al., 1997; “Childhood trauma questionnaire: A retrospective self-report: manual,” 1998). As a consequence, the classification in “exposed” vs. “unexposed” may vary even within users of the same questionnaire (see Figure 7). Moreover, the procedure of summing items to form a total or sum scores and deriving severity of exposure from this score has been criticized (Carlson et al., 2011; Lacey & Minnis, 2020). However, it has been shown that the number of different traumatic event types experienced improves the prediction of PTSD risk, whereas the additional assessment of event frequencies did not substantially enhance prediction rates further (Wilker et al., 2015). Future comparative work is necessary to shed further light on this topic. Furthermore, it might be a promising avenue for future work to take the experienced valence, controllability, or predictability (i.e., the experience) more strongly into account than the exposure itself. Together, we call for more attention to such descriptive structural properties of childhood adversity questionnaires and attempts to reduce ambiguity and hope that our systematic illustration serves as a facilitator along this path (see also BOX).

**Heterogeneity in Descriptive Structural Characteristics**

Despite all questionnaires aiming to assess childhood adversity, the specific target age group differs widely with respect to the minimum age, maximum age as well as age range. This applies to both age at exposure and age at assessment. While some questionnaires exclude late adolescence (e.g., the well established CTQ-CF, target age: 0-17 years) other questionnaires specifically focus on late adolescence and early adulthood - a developmental time well known to be characterized by substantial brain maturation processes (Giedd et al., 1999; Houston et al., 2013; Sowell et al., 1999, 2001). Relatedly, across both content analyses, more than half of all 35 questionnaires assess information on the developmental timing of adverse events and the duration or frequency of exposure (see Figure 3). This seems indeed informative as positive correlations between childhood adversity chronicity and the severity of Posttraumatic Stress Disorder (PTSD) has been reported (English et al., 2005; Jonson-Reid et al., 2012).

**Valence and Controllability**

Furthermore, although predictability has been highlighted to play a crucial role in the mechanisms linked to psychopathology according to a prominent theory on the consequences of childhood adversity (DMAP, McLaughlin et al., 2021 none of the questionnaires included here included assessments of (un)predictability. Relatedly, a single questionnaire assesses controllability of individual adversity and 8 of 35 questionnaires...
included assessments of the valence of adversity - although in some questionnaires (MACE, TAQ or ETI-SR) not for all items. These experienced characteristics of the adverse event are assumed to be associated with neurobiological changes and should be assessed in the future (Smith & Pollak, 2021). In sum, as the universally applicable childhood adversity assessment instrument yet has to be invented, our systematic and comparative overview on different childhood adversity questionnaires, their specific content and key structural characteristics (target age group, assessment of predictability, valence or controllability) will aid researchers select the most appropriate instrument for their specific purposes. In addition, we anticipate that this overview will facilitate cumulative knowledge generation and critical comparisons across results in the literature by considering the respective childhood adversity questionnaire specific advantages and disadvantages. To this end, it might be more important to synchronize screening instruments and subsequently apply in-depth instruments, to balance different demands and needs at different stages of the research process.

Limitations

Several limitations of our work deserve mentioning. First, the list of childhood adversity questionnaires used is not exhaustive, as additional questionnaires may include exposure types not covered here. This would, however, most likely increase rather than reduce heterogeneity. Second, while focusing exclusively on questionnaires aligns with previously published item-based content analyses (e.g., Karstoft and Armour, 2022 on General Trauma; Fried, 2017 on depression), content agreement between questionnaires and interview-based assessment of childhood adversity may be a valuable avenue for the future. Third, despite high inter-rater agreement in coding of items (K = .86 and K = .81), this approach cannot be fully standardized. In the future, AI tools may be a promising avenue for helping to disambiguate this task. For maximal transparency, data files and analysis code of our work are publicly available. Relatedly, assigning childhood adversity categories to items proved to be quite challenging for some items. For instance, the item “Tied up or locked in a closet.” (ETI), cannot be unambiguously categorized as either emotional or physical abuse - both of which are commonly used childhood adversity types in the literature - as it clearly contains elements of both. This example showcases a deeper challenge: the general challenge of assigning complex experiences into distinct categories. As illustrated by the example above, these are often overlapping and have hence been referred to as ‘fuzzy categories’ (Smith & Pollak, 2021, 2022). An additional criticism with respect to such subcategories is that there is in fact little evidence in the literature that specific subcategories do indeed map onto distinct, specific (neurobiological) alterations (Smith & Pollak, 2022; Young et al., 2019). In fact, it has been shown that stress-response systems are not sensitive to specific exposures, but more to individual differences and specific characteristics of experiences (Korte et al., 2005; Smith & Pollak, 2021). Furthermore, based on the available evidence (McLaughlin et al., 2021; Pollak & Smith, 2021; Smith & Pollak, 2021) it seems more promising for future research to consider the subjective experience and evaluation as well as chronicity and developmental timing of childhood adversity rather than focusing exclusively on such ‘fuzzy’ categories (Danese & Widom, 2021; Smith & Pollak, 2022).

Summary

In sum, our systematic overview of item-based content (i.e., adversity types) of questionnaires assessing childhood adversity identified substantial heterogeneity in assessment tools and operationalization as a general challenge that may in part originate from a lack of agreed upon definition of childhood adversity. Our work may serve as a starting point for moving towards “an empirical basis for classifying adversity” (Pollak & Smith, 2021). The comprehensive overview on the item-based content as well as structural characteristics of different childhood adversity questionnaires will facilitate selection of tools tailored to the specific aims of a study. In addition, it allows researchers to easily identify differences and commonalities across assessment tools which is particularly valuable when aiming to bring together results across different studies. It is important for the facilitation of cumulative knowledge generation to discuss where to go from here and how to meet the challenges and opportunities in future work and in interpreting the existing literature. To this end, we provide a detailed list of actionables (see BOX) for guidance.

We anticipate that our work will aid the improvement of comparability, replicability and cumulative knowledge generation in this societal and clinically highly relevant research field.

Box: Methods-focused Considerations and Future Directions to advance research in measuring and reporting Childhood Adversity

these considerations include general considerations as well as considerations that can be distilled directly from the present work
Selection of an childhood adversity questionnaire for a research project:

- If available, questionnaires with at least satisfactory psychometric properties according to all COSMIN criteria (Mokkink et al., 2010) should be prioritized even though we acknowledge that classical psychometric criteria may not be applicable to questionnaires assessing childhood adversity (i.e., no underlying latent construct that is to be assessed) and may not even be desirable in questionnaires aiming to assess a broad range of exposure types (e.g., Cronbach's alpha, Flake et al., 2017).

- Fit between content (i.e., adversity types) covered by a questionnaire and the research question and sample should be critically considered. For instance, the intended use of the questionnaire (such as screening or diagnosis) as well as the specific group being targeted (retrospective self-reports in adults vs. pediatric samples).

- (Additional) Assessment tools that enable a fine-grained evaluation of potentially relevant childhood adversity characteristics (i.e., onset and duration of exposure, controllability, (un)predictability) should be considered and reported (also descriptively when not of key interest to the studies aims).

- For the assessment of deprivation, it's crucial to utilize questionnaires that enable a detailed characterization of the construct rather than relying on a single item. However, it's noteworthy that most of the questionnaires examined in this analysis did not meet this criterion. Even though a study may focus on a specific type of childhood adversity, it is highly recommended to provide a nuanced screening and overview also on other childhood adversity types. This disambiguates the integration of results across studies and facilitates cumulative knowledge generation (Rösel et al., 2024).

Reporting Standards:

- Questionnaires used should be described in sufficient detail (i.e., content, target age group, subscales, types of maltreatment are assessed) and psychometric properties should be provided if applicable (however see above first bullet point).

- To facilitate cumulative knowledge generation, prevalence of adversity types (e.g., based on subscale scores and, if available, severities) in the respective sample should be characterized and reported even though this may not be the focus of a specific study (e.g., supplementary material).

Fostering reproducibility:

- Ad-hoc modifications of questionnaires should be avoided as these (may) compromise construct validity and reproducibility (Flake & Fried, 2019).

- If modifications are unavoidable, these must be reported with sufficient detail, ensuring that construct validity, and sound scientific reasoning are maintained (see Flake et al., 2017 for guidelines).

- Validated cut-offs should be adhered to and detailed cut-off information provided. Simply referring to prior publications may often lead to dead-ended reference chains. We highlight that cut-off scores may be more meaningful with respect to the experienced valence, controllability, and predictability of experiences, rather than for the number of adverse experiences.

- To facilitate cumulative knowledge generation, materials used in the assessment, such as questionnaires and interviews, should be made openly available, although copyright restrictions need to be considered.

- Generally, we call for more sharing of materials as only 26 out of 35 questionnaires were openly available. Publicly available material will facilitate rapid testing of new and existing hypotheses or psychometric validations (e.g., test-retest reliability), thus a mutual benefit of the research community, however, copyright should be considered.

Integration into the body of literature:

- To compare the results between different studies focusing on childhood adversity, the challenges of different questionnaires used and their (item-wise) content overlap as well as structural characteristics (e.g., developmental timing, experience vs. exposure, prospective vs. retrospective assessment Danese and Widom, 2020) needs to be considered even though no data may be available to allow for a direct comparison.

- Prevalences of trauma types in the respective sample, specifically the trauma load of the control group (if available in methods) should be considered.
• Exposure and experience should be distinguished in assessment and interpretation (McLaughlin et al., 2021) e.g., by using questionnaires which assess the experienced valence, controllability, and predictability of the event. While exposure directly relates to the fact of being present at a specific adverse event, experience involves subjectively perceiving it as aversive.

Author Contact
Correspondence concerning this article should be addressed to Alina Koppold, Postal address. E-mail: al.koppold@gmail.com
Alina Koppold - ORCID https://orcid.org/0000-0002-3164-3389
Julia Ruge - ORCID https://orcid.org/0000-0001-5818-5683
Tobias Hecker - ORCID https://orcid.org/0000-0001-9272-0512
Tina B. Lonsdorf - ORCID https://orcid.org/0000-0003-1501-4846

Acknowledgements
The authors thank Eiko Fried for making his R code available (Fried, 2017).

Conflict of Interest and Funding
The authors declare no competing financial interests. Financial Support This work was funded by grants by the German Research Foundation (Deutsche Forschungsgemeinschaft) to TBL: DFG LO1980/4-1 and GRK 2753/1 – Project Number 449640848, subproject A01.

Author Contributions
The authors made the following contributions. Alina Koppold: Conceptualization, Methodology, Investigation, Validation, Formal analysis, Visualization, Data Curation, Writing - review & editing; Tina B. Lonsdorf: Conceptualization, Writing - Original Draft, Supervision, Funding acquisition. Author names order connects to the contribution (AK first author, JR & TH co-authors, TBL senior author).

Open Science Practices
This article earned the Open Data, Open Materials, and Open Code badge for making the data, materials, and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References
Berkelaar, M., et al. (2022). Lpsolve: Interface to 'lp_solve' v.5.5tosolvelinear/integerprograms [https://CRAN.R-project.org/package=lpsolve].
try, 50(9), 699–704. https://doi.org/10.1016/S0006-3223(01)01167-2
Follingstad, D. R., Coyne, S., & Gambone, L. (2005). A representative measure of psychological ag-


Morrison, K. E., Stenson, A. F., Marx-Rattner, R., Carter, S., Michopoulos, V., Gillespie, C. F., Powers, A.,...


