

Validation of the “Darryl” PTSD cartoon test with abused children

Sille Schandorph Løkkegaard^{1,*} <https://orcid.org/0000-0002-0091-6508>, Camilla Jeppesen¹, <https://orcid.org/0000-0002-6312-0339>, Ask Elklit¹, <https://orcid.org/0000-0002-8469-7372>

¹Danish National Centre for Psychotraumatology, Department of Psychology, University of Southern Denmark

*Corresponding author: sschandorph@health.sdu.dk

Abstract

Background: There is a lack of reliable and valid PTSD tests for young children that cannot read or are weak readers. The semi-projective cartoon test, “Darryl”, which is read aloud, is a measure that appeals to this age group. The test has been applied in both clinical and epidemiological studies.

Objective: to validate a cartoon test, “Darryl”, for children aged 6 or older in a population of children suspectedly sexually and/or physically abused.

Methods: In the Danish Child Centres, 327 children were screened with Darryl as part of an assessment for further intervention. The Bech Youth Inventory was filled out by 113 children, and 63 caregivers filled out the Strengths & Difficulties Questionnaire. Correlations were used to study the convergent validity between the scales and subscales and effect sizes were estimated. Reliability of the scales was investigated using Cronbach’s alpha.

Results: Following the DSM-IV, 55.7% of the children (n = 182) had a possible PTSD diagnosis. More girls (n = 110, 62.9 %) than boys (n = 72, 47.4%) had PTSD. A total of 21.7 % (n = 71) had subclinical PTSD, lacking only one symptom of the full diagnosis. There was no significant difference in PTSD regarding physical or sexual abuse.

Clinical significance: The test allows clinicians in the pediatric field to screen for possible PTSD in a population where systematic self-report data have paramount importance.

Conclusions: Darryl appears to be a valid and reliable test for screening young children who have been physically or sexually abused. The test is helpful for clinician working with young children to identify those have developed trauma symptoms to secure them early treatment.

Keywords: young children, trauma, abuse, cartoon test, validation

Validation of the “Darryl” PTSD cartoon test with abused children

Research on trauma exposure and development of PTSD in children has increased. Studies have shown that most children are exposed to one or more potential traumatic events or childhood adversities during their lives (1-3). However, studies have also shown that a minor group of children are at risk of being exposed to several traumas and also certain types of traumas such as sexual, physical, and emotional abuse (4,5) and have high risks of developing PTSD later in life (6) and experiencing a range of other negative social, mental, and physical consequences (7). Similarly to studies of adult trauma survivors, a recent study found that a 78% co-

morbidity between PTSD and other axis I disorders in a group of 80 traumatized and abused Danish preschoolers (8). The disorders included oppositional defiant disorder, separation anxiety disorder, and major depression as the most common. The endorsement of PTSD in this group was 46.3%.

Therefore, it is of most importance that clinicians can identify traumatized children as early as possible to allocate appropriate treatment. However, the younger the children, the more difficult it becomes to reliably assess reactions and symptoms following trauma exposure by child reports because children are still developing their cognitive, emotional, and behavioral skills. Many abused children have shown cognitive deficits. Cicchetti & Rogosch (9) found that

abused children demonstrated inhibitions in talking about themselves and discussing negative emotions and thoughts. Also, many abused children may be less aware of negative feelings, more emotionally on guard, or have a great need for social approval (10). Post-traumatic reactions have included lower IQ and reading achievement (11), delayed language and poorer performance at school (12). Furthermore, children with PTSD have significant impairments in attention, abstract reasoning, and executive functioning compared to healthy controls (13).

This impacts how researchers and clinicians can ask questions in interviews or use questionnaires about some of the rather abstract symptoms of trauma, how children understand these questions, the children's insight into own bodily, emotional, cognitive, and behavioral reactions, and the children's ability to reliably answer such questions orally or by reading and checking scales or boxes in a written questionnaire. Pictorial cartoon tests accommodate some of these difficulties.

Two other cartoon tests for children's PTSD symptoms have been reported with validation evidence. The Cameron Complex Trauma Interview (CCTI) has the advantage of determining the potential traumatic event from the ten-item Adverse Childhood Experiences list introduced in the beginning and using a gender-neutral puppy. Eight psychological domains (attachment, biology, affect regulation, dissociation, behavioral control, cognition, self-concept, PTSD symptoms) are investigated by means of 21 symptoms on a 4-point visual Likert scale (with 'no answer' added as a 5th step) where the level of distress in Cameron's face is used as indicator (14). This test has been used in one published study of 21 children collected through a snowball procedure. There is no reported investigation of the validity of the psychological domain items but clinicians' ratings of using the test were positive. The clinicians were recommended to seek typical 'who, what, where, when, how' details for each of traumatic experiences and told "that at this point you do not need to inquire about emotional reactions". The ACE part and the symptom parts correlated positively ($r=.47$), and both parts correlated well with the UCLA PTSD Reaction Index-5. The sample was small, and the selecting procedure biased. There have been no further studies using the test since the original study (15).

Another pictorial trauma test for young children is the Angie/Andy cartoon test (ACTS) that has been more thoroughly investigated. Four groups of children (intra-, extra-familial, combined, and non-trauma; $n = 208$) were compared on 22 PTSD items and 65 Complex PTSD items corresponding to the tentative DESNOS diagnosis in DSM-IV (16). There is no anchoring event; the index trauma must be

determined before the start of the test. Angie/Andy was compared to a test, Child Rating Scales of Exposure to Interpersonal Abuse developed especially for this study as a research tool; the items being taken from the Conflict Tactics Scales (17) covering physical abuse, witnessing family and community violence. In addition, sexual violence was also covered. The article does not report the number of exposure items within the various groups of violence, nor is there any description of which items that has been assigned with a double weight. A parent rating scale was used to examine the concurrent validity of the test. The internal consistency was high, and the construct and concurrent validity was promising (18). The ACTS is no longer in print. For the time being due to lack of availability, lack of investigated construct validity, and procedural demands on extensive questioning about trauma exposure details, we do not consider the abovementioned tests viable alternatives.

A third cartoon test is "Darryl", which was developed to examine cardinal symptoms of DSM-IV PTSD in school-aged children exposed to community violence (19). The Darryl includes both written questions of PTSD symptoms, which is read aloud by the interviewer to the child and accompanying pictures, which illustrate the preadolescent character, "Darryl", who experiences each symptom after exposure to community violence. The child is then asked to rate how often they feel like Darryl. By using both auditory and visual stimuli during the assessment, the child's attention and focus is maintained during a demanding task of talking about symptoms of trauma. The pictures assist in recognizing sensory elements from the child's own abuse. The combination of pictures and a very short text read aloud followed by a simple question intends to provide a gentle, safe space for children to reveal their inner worlds. They can identify with another child so that they do not feel so alone with their feelings and symptoms. They can also point to the symptoms on the thermometers, which is a safe distance from themselves, as they do not have to verbalize. This approach is indirect, free of blaming, and non-threatening. The scoring choices are also adapted to a young child as they consist of three thermometers, one empty, one half full, and one filled, indicating whether the child 'never', 'some of the time' or 'a lot of the time' can relate to the symptom described in the picture (20).

Darryl was originally validated with a non-clinical sample of 110 American children aged 7-9 exposed to community violence (19). Internal consistency was found to be excellent for the overall scale of PTSD ($\alpha = .92$) and acceptable to good for the subscales of reexperiencing, avoidance, and hyperarousal,

respectively ($\alpha = .78$, $\alpha = .83$ and $\alpha = .80$). Child reports of summed exposure scores were significantly associated with summed symptom scores on Darryl ($r_s = 0.48$, $p < .001$), as were adult reported exposure scores and the Darryl summed exposure scores ($r_s = 0.29$, $p < .006$). A second study with a clinical sample of 49 children and adolescents also found excellent overall internal consistency ($\alpha = .90$) and adequate to acceptable internal consistencies for the three subscales ($\alpha = .72$, $\alpha = .69$ and $\alpha = .65$). Also, the overall symptom score on Darryl was significantly correlated with the overall symptom score on the Child's Post-traumatic Stress Disorder Reaction Index (CPTSD-RI, $r = 0.64$, $p < 0.001$) (21).

In Denmark, the original version of Darryl has first been translated and adapted for use in assessment of children who experienced a firework factory explosion in a residential area (22-23). In addition, five different versions covering other types of traumas have been developed, including exposure to physical abuse, sexual abuse, war, disease, and death. Depending on the reason for referral to assessment, the corresponding version of the Darryl will be chosen. However, the different versions also call for studies examining the external validity of each version of the Darryl. Currently, the version used for exposure to a disaster (explosion) and exposure to violence have been validated with a sample of 430 children, aged 6-18, who were affected by the fireworks explosion (22) and a clinical sample of 65 children, aged 7-17 years, exposed to physical abuse (20). Both studies found good internal consistency for the overall scale of the Darryl ($\alpha = .89$, (22); $\alpha = .88$, (20). However, Elklit and colleagues (22) reported a Cronbach's α of .59 for the subscale of hyperarousal, acceptable internal consistency for the subscale of reexperiencing ($\alpha = .71$) and less than adequate internal consistency for the subscale of avoidance ($\alpha = .62$). Internal consistencies for the version on physical abuse were adequate to acceptable ($\alpha = .78$ for reexperiencing, $\alpha = .69$ for avoidance, and $\alpha = .75$ for hyperarousal) (20). Validity of the disaster version was examined using Darryl results from 215 children aged 12-18 years, where scores on Darryl and the Harvard Trauma Questionnaire (HTQ) and on Darryl and the Trauma Symptom Checklist (TSC) were both strongly and significantly correlated ($r = .76$, $p < .01$ and $r = .74$, $p < .01$, respectively) (22).

When the Darryl version for physical abuse was compared to the child and adolescent version of the Clinician Administered PTSD Scale (CAPS-CA) (24), scores were also strongly and significantly correlated ($r = .73$, $p < 0.001$) (20). The present study seeks to add to existing studies on the Darryl by examining the validity of two versions of the Darryl which are

used in the Danish Children Centers (DCC), which assess children who are referred to the centers based on suspicion to exposure to physical and/or sexual child abuse (25).

Aims of the present study

The present study aims to examine reliability and validity of two versions of the Darryl cartoon test for use after exposure to physical abuse or sexual abuse. The first objective is to report PTSD symptomatology in children aged 6-17 years, who had been referred for psychological assessment due to suspected exposure to physical and/or sexual abuse. The second objective is to examine the reliability of the Darryl scale and subscales, and the third objective is to examine convergent validity of the Darryl when compared to another child-reported measure, the Beck Youth Inventories (BYI-II) (26-27) and when compared to the Strengths and Difficulties Questionnaire (SDQ) (28-29) as reported by child caregivers. Thereby, this study provides further validation of the Darryl cartoon test and is the first study to examine reliability and validity of the version for assessment of children exposed to sexual abuse.

Method

Participants and procedures

A total of 327 children (152 boys and 175 girls) from all the DCCs were included in this study. The DCCs are five governmentally initiated centers that offer specialized support and in collaboration with municipalities coordinate cases of suspected child abuse. When two or more public sectors (e.g., municipality, police, or hospital) are involved in a case of suspected child abuse, the local DCC are mandated to coordinate the cross-professional work on the case. The DCCs are placed in each of the five regions of Denmark. The DCCs are child-centered which means that the child goes to the DCC and all professionals such as caseworkers, police, and prosecutors come to the DCC to work on the case. Also, by law the centers have to be child friendly in their interiors. Furthermore, the DCCs offers a specialized trauma-focused psychosocial screening of the child's need for support and treatment no matter the turn-out of the legal case.

A team of one psychologist and one social worker conducts the screening of the child. The screening is conducted during up to 4 meetings with the child and a caregiver. All DCCs use a national assessment battery which include standardized and validated assessment tools for children aged 1-17 years. It is up to the professionals, who conduct the screening, which measure(s) to use depending on the case content and the child's age, cognitive skills, and willingness to participate and disclose symptoms and

mental state (25). The Darryl cartoon test is part of the national assessment battery and available in two versions in the DCCs: one for use after exposure to physical abuse and one for use after exposure to sexual abuse. In the DCCs, the Darryl test is used in cases where the children to some extent are able to talk about the abuse on which they have been referred to the DCCs.

Data were collected between May 2018 and December 2020. All included children answered a Darryl cartoon test. A subsample of 113 children also filled out the BYI-II and caregivers for 63 children filled out the SDQ caregiver version. The study was approved by RIO, the Data Protection Agency of the University of Southern Denmark.

The children were between the ages of 6 to 17 years. Mean age of was 9.9 years ($SD = 2.18$). The distribution by age groups was: 146 children were between 6 and 9 years, 144 children were between 10 and 12 years and 37 children were between 13 and 17 years of age.

Measures

Physical and sexual abuse

For each case in the DCCs, the employees at the DCC register whether the child is referred to the DCC due to suspicion of exposure to physical and/or sexual abuse. These data and the demographic data are sent for administrative purposes to the National Board of Social Affairs who monitors the field.

The Darryl cartoon test

“Darryl” is a cartoon-based assessment tool designed to measure symptoms of possible PTSD according to DSM-IV among school-aged children (19). The test consists of cartoon pictures illustrating the boy Darryl (in Danish “Thomas”) and is designed to illustrate the boy in different situations where emotional, behavioral, or cognitive symptoms of PTSD occur. Each cartoon depicts a different PTSD symptom, and an interviewer reads aloud the corresponding short script describing the content of the cartoon. The child is asked to respond to the script using one of the three response choices “never”, “some of the time” or “a lot of the time” to describe to what extent the child feels like Darryl. The responses are assigned a score of 0, 1, or 2, respectively (8). The tests used in this study are the versions designed to reflect one of the two traumas: sexual abuse or physical abuse. The Danish version of the tests consist of 19 items which measure symptoms of possible PTSD which gives a possible range of scores between 0-38. Seven items cover symptoms of re-experiencing, seven items cover symptoms of avoidance, and five items cover symptoms of hyperarousal (22). Two additional

items concern somatic symptoms (headache and stomachache) that are common trauma symptoms in children. Following the DSM-IV (16), diagnostic criteria for PTSD were met if participants endorsed at least one symptom of re-experiencing, at least three symptoms of avoidance and at least two symptoms of hyperarousal. The symptom criteria were met if the response was “Some of the time” or “A lot of the time”.

The Beck Youth Inventories, second edition (BYI-II)

The BYI-II is a self-report measure which assess the child’s perceptions of depression, anxiety, anger, disruptive behavior, and self-concept. The Beck Depression Inventory (BDI), Anxiety Inventory (BAI), Anger Inventory (BANI), Disruptive Behavior Inventory (BDBI), and Self-concept Inventory (BSCI), each contains 20 statements, and the child is responding to the questionnaire to indicate the extent to which they think that a sentence characterizes them. They answer on a four-point Likert-scale (0 = never, 1 = sometimes, 2 = often and 3 = always). The five inventories are scored by adding up the 20 ratings. Each inventory can be used separately. The BYI-II can be filled out by children 7 years or older (26). The BYI-II is validated with Danish samples and shows acceptable internal consistency and test-retest reliability (27). Internal consistencies for the scales were good to excellent (BDI $\alpha = .95$, BAI $\alpha = .92$, BANI $\alpha = .94$, BDBI $\alpha = .89$, BSCI $\alpha = .88$) (27).

The Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a brief behavioral screening questionnaire containing 25 items about the child’s strengths and difficulties (28). The questionnaire exists in three versions: a self-report version for use for children aged 11 years and older, a parent report version, and a teacher/pedagogue report version. In the present study, the parent report version was used. The answers to the questions are scored on a three-point Likert-scale (0 = not true, 1 = somewhat true and 2 = certainly true). The items of the questionnaire are divided into five subscales: a hyperactivity scale, emotional symptoms scale, conduct problems scale, peer problems scale, and a prosocial scale. The first four scales are summed up into a Total Difficulties scale which has a maximum score of 40 (28). The SDQ shows good psychometric value and has been validated with Danish children (29). Internal consistency for the total scale was good ($\alpha = .82$) and internal consistencies for four subscales were adequate to acceptable (hyperactivity scale $\alpha = .79$, conduct problems scale $\alpha = .69$, peer problems scale $\alpha = .72$, and a prosocial scale $\alpha = .62$), however lower for the emotional symptoms scale with a Cronbach’s $\alpha = .58$.

Statistical analyses

Statistical analysis included descriptive statistics, computing Cronbach's alpha for examining internal consistency, and correlation analysis between the Darryl scales, the scales of the BYI-II, and the SDQ parent version using Pearson's r . Bivariate correlations were applied to measure the relationship between exposure to abuse types (physical violence, sexual violence, polyvictimization) and mental health outcomes (i.e., PTSD-symptoms, and trauma-related symptoms, e.g., anxiety, depression, anger). To assess the size of a correlation, the criteria proposed by Cohen (30) were used where $r = 0.1$ constitutes a small correlation, $r = 0.3$ constitutes a medium correlation, and $r = 0.5$ constitutes a large correlation. To investigate relationships between abuse types, genders, age groups, and symptoms of psychopathology, t tests were conducted. Analysis of variance (ANOVA) is used to analyze the differences among means of PTSD symptomatology in the three abuse types as we expect the combined type to have higher levels. A post hoc test, Tukey-B was used to assess significant differences between the three types. Cohen's d was calculated for significant findings, where a Cohen's d value of 0.2 constitutes a small effect, 0.5 a moderate and 0.8 a large effect.

Results

Of the 327 children who filled out the Darryl, a total of 238 children (126 boys and 112 girls) had been referred to a DCC on suspicion for exposure to physical abuse, while 58 children (15 boys and 43 girls) had been referred due to suspected exposure to sexual abuse. A minor group of 19 children (8 boys and 11 girls) were referred to the DCCs for suspected exposure to both physical and sexual abuse. For another group of 12 children, there is no report of the type of abuse. Table 1 shows the distribution of referral reason across gender.

TABLE 1. Referral reasons across gender.

	Total		Male		Female	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Sexual Abuse	58	17.7	15	10.1	43	25.9
Physical abuse	238	72.8	126	84.5	112	67.5
Polyvictimization	19	5.8	8	5.4	11	6.6
Missing	12	3.7				
Total	327	100	149	100	166	100

Note: $\chi^2(2) = 13.938; p < .001$

PTSD symptomatology

Depending on the referral reason, the corresponding version of the Darryl had been used (i.e., for physical or sexual abuse). Of the 19 children who had been referred to the DCCs for suspected exposure of both physical and sexual abuse, nine children had answered the Darryl version for physical abuse and

ten children had answered the version for sexual abuse.

For the entire sample, the children reported between zero and 19 symptoms on the Darryl, with an average number of symptoms reported at 9.4 (SD = 4.8). A total of 25% of the children reported five or less symptoms, 51% of the children reported between six and 13 symptoms, and 24% of the children reported 14 or more symptoms. Table 2 shows the frequencies of the children experiencing the different symptoms. The most frequently experienced symptom on the Darryl was avoidance of thoughts and feelings, which was reported by 80.4% of the children. More than 3/4 of the children (76.8%) also experienced intrusive recollections. On the total scale of the Darryl, scores ranged between 0 and 38 with a mean of 12.9 (SD = 7.7). An independent-samples t -test was conducted to compare the Darryl PTSD symptoms for males and females (Table 2). There were significant differences in scores on 18 out of 21 symptoms (including one psychosomatic symptom) with girls scoring higher on 16 out of 18 symptoms. The effect sizes were small to medium (Cohen's d .018 - .428).

Following the DSM-IV criteria more than half of the children (55.7%, $n = 182$) had possible PTSD. Significantly more girls ($n = 110$, 62.9 %) than boys ($n = 72$, 47.4%) had PTSD ($\chi^2(1, 327) = 7.91, p = .005$). There was no significant difference in possible PTSD regarding referral reason ($\chi^2(1, 315) = .06, p = .808$). In addition, a total of 21.7 % ($n = 71$) had subclinical PTSD, meaning that they lacked only one symptom to live up to a full DSM-IV PTSD diagnosis based on the Darryl.

Reliability

Table 3 shows the overall internal consistencies of the Darryl for the two versions used after sexual abuse and physical abuse. The overall internal consistency of the two versions of the Darryl were excellent. The subscales for the version for sexual abuse also showed good to excellent reliability. The subscales for the version for physical abuse showed adequate to good reliability.

Criterion validity

A One Way ANOVA was performed with between abuse types (physical = 1, sexual = 2, and combined = 3), the Darryl total score and the three subscores (re-experiencing, avoidance, and hypervigilance). The F -values (2,324) 3.82, 3.17, 3.10, and 3.38; the p 's .023, .044, .046 and .035, respectively. A post-hoc Tukey Y-B analysis showed that for total scale: $3 > 1$; the re-experiencing: $3 > 1$, and for the avoidant subscale: $3 > 1, 2$). For the hypervigilant scale, there was no significant difference in scores between the three groups (Table 3).

TABLE 2. Percentage frequency and mean scores of PTSD symptom items and somatization items in the Darryl ($n = 327$).

Score	Frequency (%)			Boys ($n = 152$)		Girls ($n = 175$)		t	p	Cohen's d
	Never 0	Some of the time 1	Most of the time 2	Mean	SD	Mean	SD			
<i>Re-experiencing symptoms</i>										
Intrusive recollections	23.2	46.2	30.6	.969	.728	1.149	.727	-2.004	.046	.222
Play re-enactment of the trauma	84.7	11.9	3.4	.230	.521	.149	.416	1.576	.116	.175
Nightmares in general	39.4	42.2	18.3	.625	.708	.931	.724	-3.864	<.001	.428
Nightmares about the trauma	57.8	32.1	10.1	.415	.624	.617	.700	-2.744	.006	.304
Acting/feeling as if the event is recurring	51.1	34.6	14.4	.553	.689	.703	.745	-1.894	.059	.209
Psychological distress at reminders	56.6	28.4	15.0	.467	.699	.686	.757	-2.713	.007	.299
Physiological distress at reminders	56.0	31.2	12.8	.441	.688	.680	.712	-3.087	.002	.341
<i>Avoidance symptoms</i>										
Avoidance of thoughts and feelings	19.6	37.9	42.5	.1237	.778	1.223	.736	.166	.868	.018
Avoidance of people and activities	48.3	31.8	19.9	.651	.783	.771	.769	-1.395	.164	.155
Inability to recall important aspects of the event	38.5	38.8	22.6	.868	.795	.817	.743	.600	.549	.067
Loss of interest in usual activities	63.3	26.9	9.8	.309	.612	.600	.687	-4.028	<.001	.445
Detachment/social withdrawal	67.6	22.6	9.8	.290	.583	.709	.054	-3.466	<.001	.379
Restricted range of affect	61.2	29.4	9.5	.467	.690	.642	.049	-.406	.685	.045
Sense of foreshortened future	60.2	25.7	14.1	.480	.709	.744	.056	-1.346	.179	.149
<i>Hyperarousal symptoms</i>										
Difficulty initiating and/or sustaining sleep	39.4	29.7	30.9	.783	.797	.854	.065	2.690	.008	.297
Irritability	45.6	33.3	21.1	.697	.789	.771	.058	-1.252	.212	.139
Decreased concentration	42.8	37.3	19.9	.625	.744	.751	.057	-3.284	.001	.364
Hypervigilance	46.8	33.9	19.3	.605	.756	.761	.056	-2.655	.008	.294
Exaggerated startle response	55.4	26.0	18.7	.507	.737	.800	.060	-2.778	.006	.306
<i>Somatization</i>										
Stomachache	44.3	41.0	14.7	.579	.646	.746	.056	-3.020	.003	.332
Headache	53.8	33.0	13.1	.533	.699	.719	.054	-1.436	.152	.159

TABLE 3. Internal consistencies (Cronbach's α) for the total scale and subscales of the Darryl for the version for sexual abuse and the version for physical abuse. Also, means and SD on the Darryl scales for the two abuse types and combined abuse.

Darryl version	Sexual abuse α	Physical abuse α	Physical abuse Mean (SD)	Sexual abuse Mean (SD)	Combined abuse Mean (SD)
Total scale	.90	.87	12.260 (7.443)	13.850 (8.102)	17.50 (9.196)
Re-experiencing subscale	.86	.77	4.153 (3.106)	4.663 (3.295)	6.333 (4.097)
Avoidance subscale	.77	.69	4.549 (2.979)	4.813 (3.171)	6.750 (2.927)
Hyperarousal subscale	.88	.67	3.557 (2.647)	4,375 (3.070)	4.667 (3.172)

TABLE 4. Pearson's r correlations between the total scale and subscales of the Darryl test and subscales of Beck Youth Inventory.

	PTSD Total		Re-experiencing		Avoidance		Hyperarousal	
	r	p	r	p	r	p	r	p
BYI-II Depression ($n = 112$)	.526	<.001*	.313	<.001*	.550	<.001*	.502	<.001*
BYI-II Anxiety ($n = 113$)	.495	<.001*	.381	<.001*	.450	<.001*	.440	<.001*
BYI-II Anger ($n = 109$)	.496	<.001*	.318	<.001*	.482	<.001*	.486	<.001*
BYI-II Disruptive behavior ($n = 104$)	.241	.012*	.158	.104	.285	.003*	.188	.057
BYI-II Self-concept ($n = 112$)	-.203	.032*	-.090	.348	-.260	.006*	-.176	.064

Note. BYI-II = Beck Youth Inventory, 2nd Edition. *significant at the 0.05 level (2-tailed).

TABLE 5. Pearson's *r* correlations between total scale and subscales of the Darryl test and the SDQ Total Difficulties scale and subscales parent report version (*n* = 63).

	PTSD Total		Re-experiencing		Avoidance		Hyperarousal	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
SDQ Total Difficulties	.196	.013*	.091	.476	.200	.117	.234	.084
Emotional problems	.284	.024*	.188	.141	.252	.046*	.317	.011*
Conduct problems	.148	.248	.116	.391	.107	.402	.177	.164
Hyperactivity	.125	.330	.050	.695	.161	.208	.122	.342
Peer Problems	.001	.996	-.084	.513	.039	.760	.057	.658
Prosocial behavior	.061	.633	.078	.539	-.012	.924	.098	.446
Impact scale (<i>n</i> = 59)	.241	.066	.070	.601	.275	.035*	.301	.020*

Note. SDQ = Strengths and Difficulties Questionnaire. *significant at the 0.05 level (2-tailed).

Convergent validity

To examine the convergent validity of Darryl, scores on scales of the Darryl were compared to the scores on the scales of the self-report BYI-II and the parent report version of the SDQ. Scores on the BYI-II and the SDQ are shown in Table 4 and Table 5. Overall, three symptom scales of the BYI-II (depression, anxiety, and anger) correlated highly with all three subscales and with the total scale of Darryl. However, there is some preliminary evidence of criterion validity specific to PTSD based on the lower correlations between the re-experiencing score and the other internalizing symptom scores (~.30 vs. .45-.55 for the more non-specific PTSD avoidance and arousal symptoms).

The disruptive behavior subscale correlated with the avoidance subscale and the total Darryl scale. A positive self-concept was associated with little avoidance and a lower total Darryl score.

The parental or caregiver SDQ version only had few significant correlations with the Darryl scales. The SDQ emotional subscale was positively associated with the avoidance and the hyperarousal subscales, and so was the impact scale that indicate the overall impact of the symptoms on the function of the child and the family. The conduct problems scale, the hyperactivity scale, and the peer problems scale were not even close to be associated with the PTSD subscales and nor was the prosocial behavior scale.

Discussion

This study set out to examine reliability and validity of the two versions of the Darryl cartoon test for use after exposure to physical abuse and sexual abuse. The study is the first to examine reliability and validity of the version for use after sexual abuse. Both versions performed well with a satisfactory reliability. The two abuse groups did not differ significantly on any of the Darryl scales but the group who had been exposed to both types of abuse (polyvictimization) scored remarkably higher on all scales except the hypervigilant scale. This finding supports the criterion validity of Darryl scale, as polyvictimization generally in a dose-response way relates to higher

levels of psychopathology and specifically to PTSD (31).

The convergent validity with the BYI-II was convincing with each of the five subscales correlating with the Darryl scales, strongest with the avoidance subscale and the total PTSD scale. The lesser associations with the disruptive and the self-concept subscales might reflect that the assessments are conducted shortly after the disclosure of the abuse and therefore disruptive behavior has not yet become as pronounced as it might after longer time if no preventive steps have been taken. At the same time, disruptive behavior is less common as depression and anxiety. The same more slowly developing deterioration can be said of the self-concept or self-worth which could be more devastated as time go by without any help.

The general lack of association between Darryl and the parent version of the SDQ is understandable when one considers that often the parent is not fully aware of the younger child's problems due to cognitive and expressive limitations of the younger child. This caretaker underestimation can be conceived as a part of reciprocal protection in the family, which is very common due to the shame and taboos associated with child abuse. However, two subscales, the emotional and the impact scales had significant correlations with the avoidance and total PTSD scales. So, when the child reports avoidant symptoms and behavior, it seems related to the emotional problems registered by the parent and this impacts both the child and the family. Both with the BYI-II and the SDQ, the Darryl subscale of Avoidance seems to play a pivotal role linked to high scores on emotional problems and overall impact as well as more disruptive behavior and low self-worth. Although the analyses support both the reliability and the validity of the two Darryl versions for sexual and psychical abuse, there are several limitations to be aware of. First, recently abused children who are police interrogated and whose disclosure has caused alarm in their families and social network, are in an acute crisis. They must be treated gently and with compassion meaning that one should not exposed

them to more challenges than absolutely necessary. This means that a 'golden, perfect' test situation with intensive and extensive interviews, repeated testing with the same instrument, or with several similar instruments is not possible at the DCCs. Therefore, we received the test data that the responsible clinical psychologist had chosen for a particular child without questioning the choices made.

Secondly, there is a dearth of trauma tests for young children. The CAPS-CA can in our experience be extremely stressful and potentially highly distressing due to its length and details for many seven years old children who are in a crisis but okay for most adolescents who are not in an acute crisis (20). Several questionnaires exist that are used with seven years old children, but very few, if any, build on a thorough examination of the child's understanding of critical diagnostic trauma symptoms, the child's understanding of time and grasp of time periods, or the child's ability of express the degrees of symptom severity or frequency on a Likert scale with several or many steps.

The lack of a robust and simple validated Danish PTSD measure for very young children to assess the concurrent validity is a limitation in the current study. Future validation studies will have to rely on a clinical assessment of PTSD. This will take extensive training, more extensive parent or caregiver involvement, traumas that are not embedded in an acute or dramatic situation with possible dissolution of the family, and extended time for assessment compared to the present study.

The semi-projective methodology of the cartoon test with identification with a troubled child in a specific abuse or trauma context appeals to many young children as meaningful, and the translation from the plight of the figure to the child itself through pointing at a thermometer seems to work well even for very young children (six-seven years). The Darryl also seems to work well with older children and even adolescents in this specific assessment situation at the DCCs. The sample is relatively large and includes approximately equal numbers of each sex who have been victims of sexual or physical abuse who have been police interrogated and assessed by trained psychologist according to Danish law. The reliability for both versions of abuse were good and the convergent validity based on the BYI-II and the SDQ for the older and the younger age groups, respectively, was also good and meaningful.

For this young group of abused children, the development of a suitable measure with adequate psychometric properties is of paramount importance. Accurate assessment leads to the correct diagnosis and facilitates a focused clinical intervention that alleviates the symptom burden of the child. With the changes in diagnostic systems, a

future task will be to make a validation of the various Darryl versions according to the ICD-11 (32) and the DSM-5 (33). Besides the change of the pictures and the text, the collection of data from this population of abused children will take some time because data collection is embedded in a clinical practice where also other authorities (e.g., the police, the courts, and the municipalities) are stakeholders. However, the data collection for the ICD-11 version of Darryl, that encompasses complex PTSD, is on its way. The ultimate quality of the Darryl is that it gives the young child a voice so that we can understand and react according to the symptoms and suffering the abused child reports.

Disclosures

The authors have no competing interests to declare

Funding

This project was supported by the Victims' Foundation (18-910 00085) and the Ole Kirk's Foundation.

References

1. Copeland WE, McGinnis EW. Epidemiology of Trauma and PTSD in Childhood and Adolescence (pp. 76-97). In M. J. Friedman, P. P. Schnurr, & T. M.Keane (Eds.): Handbook of PTSD. New York: Guilford; 2021.
2. Karsberg S, Lasgaard M, Elklit A. Victimization and PTSD in a Greenlandic youth sample. *International Journal of Circumpolar Health*. 2012; 71:18378. <https://doi.org/10.3402/ijch.v71i0.18378>
3. McLaughlin KA, Greif Green J, Gruber M J, Sampson NA, Zaslavsky AM, Kessler RC. Childhood adversities and first onset of psychiatric disorders in a national sample of US adolescents. *Arch Gen Psychiatry*. 2012; 69(11); 1151-1160. <https://doi.org/10.1001/archgenpsychiatry.2011.2277>
4. Armour C, Elklit A, Christoffersen MN. A latent class analysis of childhood maltreatment: Identifying abuse typologies. *Journal of Loss and Trauma*. 2014; 19(1); 23-39. <https://doi.org/10.1080/15325024.2012.734205>
5. Shevlin M, Elklit A. A latent class analysis of adolescent adverse life events based on a Danish national youth probability sample. *Nordic Journal of Psychiatry*. 2008; 62(3); 218-224. <https://doi.org/10.1080/08039480801983992>
6. Murphy S, Shevlin M, Armour C, Elklit A, Christoffersen MN. Childhood adversity and PTSD experiences: Testing a Multiple Mediator Model. *Traumatology*. 2014; 20(3); 225-231. <https://doi.org/10.1037/h0099838>
7. Elklit A, Murphy, S. The nature and extent of child maltreatment in Denmark: Using self-report and administrative data from the 1984 birth cohort. *Child Abuse & Negl*. 2022. Doi: 10.1016/j.chiabu.2022.105786
8. Løkkegaard SS, Egebæk SAB, Elklit A. Are trauma and post-traumatic stress disorder connected to psychiatric comorbidity in Danish pre-schoolers? *Journal of Child and Adolescents Trauma*. 2017;10(4); 353-361. Doi: 10.1007/s40653-017-0146-z
9. Cichetti D, Rogosch FA. The toll of child maltreatment on the developing child. In SJ Kaplan & D Pelcovitz (Eds), *Child and Adolescent psychiatric clinic of North America*. Philadelphia: W.B. Sanders; 1994.

10. Shapiro JP, Leifer M, Martone MW, Kassem, L. Multimethod assessment of depression in sexually abused girls. *Journal of Personality Assessment*. 1990; 55; 234-238.
11. Delany-Black V, Covington C, Ondersma SJ, Nordstrom-Klee BA, Templin TM, Ager J, et al. Violence exposure, trauma, and IQ and/or reading deficits among urban children. *Arch Pediatr Adolesc Med*. 2002; 156(3); 280-285.
12. Shonk SM, Cicchetti D. Maltreatment, competency deficits, and risk for academic and behavioral maladjustment. *Developmental Psychology*. 2001; 68; 670-683.
13. Beers SR, de Bellis MD. Outcomes of child abuse. *Neurosurg Clin North Amer*. 2002; 13; 235-241.
14. King JA. Getting the picture: A cartoon-based assessment tool for complex trauma in children. [https:// repository.upenn.edu/dissertations_sp2/68](https://repository.upenn.edu/dissertations_sp2/68). 2015
15. King JA, Solomon P, & Ford JD. The Cameron complex trauma interview (CCTI): Development, psychometric properties, and clinical utility. *Psychological Trauma*. 2017; 9; 18-22.
16. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 4th ed. Washington, DC: Author; 1994.
17. Strauss MA, Gelles RJ. Measuring intrafamilial conflict and violence. The Conflict Tactics (CT) Scales. *Journal of Marriage and Family*. 1977; 41; 75-88.
18. Praver F, DiGiuseppe R, Pelcovitz D, Mandel FS, Gaines R. A preliminary study of a cartoon measure for children's reactions to chronic trauma. *Child Maltreatment*. 2000; 5; 273-285.
19. Neugebauer R, Wasserman GA, Fisher PW, Kline J, Geller PA, Miller LS. Darryl, a cartoon-based measure of cardinal posttraumatic stress symptoms in school-age children. *Am J Public Health*. 1999; 89(5); 758-761.
20. Løkkegaard SS, Rønholt S, Karsberg S, Elklit A. Validation of the PTSD screening cartoon test "Darryl" in a Danish clinical sample of children and adolescents. *International Journal of Methods in Psychiatric Research*. 2017; 26(1). <https://doi.org/10.1002/mpr.1514>
21. Geller PA, Neugebauer R, Possemato AK, Walter P, Dummit ES, Silva RR. Psychometric properties of Darryl, a cartoon based measure to assess community violence-related PTSD in children. *Psychiatr Q*. 2007; 78(2); 157-168. <https://doi.org/10.1007/s11126-006-9035-6>
22. Elklit A, Nielsen LH, Lasgaard M, Duch CA. Cartoon-Based Measure of PTSD Symptomatology in Children Exposed to a Disaster. *Journal of Loss & Trauma*. 2013; 18(1); 54-63. <https://doi.org/10.1080/15325024.2012.679125>
23. Elmoose M, Duch C, Elklit A. Children's coping styles and trauma symptoms after an explosion disaster. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*. 2016; 4(3); 132-140.
24. Nader KO, Kriegler JA, Blake DD, Pynoos RS, Newman E, Weather FW. Clinician administered PTSD scale, child and adolescent version. White River Junction, VT: National Center for PTSD. 1996; 156.
25. Løkkegaard SS, Vang ML, Elklit A. (submitted). When Multidisciplinary Clinical Practice and Research Meet: Quality Development in the Danish Children Centres.
26. Beck JS, Beck AT, Beck JJB. Youth inventories of emotional & social impairment: Depression inventory for youth, anxiety inventory for youth, anger inventory for youth, disruptive behavior inventory for youth, self-concept inventory for youth: Manual. Psychological Corporation; 2001.
27. Thastum M, Ravn K, Sommer S, Trillingsgaard A. Reliability, validity and normative data for the Danish Beck Youth Inventories. *Scand J Psychol*. 2009; 50(1); 47-54. <https://doi.org/10.1111/j.1467-9450.2008.00690.x>
28. Goodman R. The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*. 1997; 38(5); 581-586.
29. Niclasen J, Teasdale TW, Andersen AMN, Skovgaard AM, Elberling H, Obel C. Psychometric properties of the Danish Strength and Difficulties Questionnaire: The SDQ assessed for more than 70,000 raters in four different cohorts. *Plos One*. 2012; 7(2). <https://doi.org/ARTN e32025> 10.1371/journal.pone.0032025
30. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Lawrence Erlbaum; 1988.
31. Finkelhor D, Ormrod RK, Turner HA. Poly-victimization: a neglected component in child victimization. *Child Abuse Negl*. 2007; 31(1);7-26. doi: 10.1016/j.chiabu.2006.06.008.
32. World Health Organization. *International classification of diseases for mortality and morbidity statistics (11th revision)*. 2018.
33. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR)*. Washington, DC: Author; 2022.