



Published in final edited form as:

Neuropsychology. 2015 January ; 29(1): 31–44. doi:10.1037/neu0000095.

The Childhood Cancer Survivor Study-Neurocognitive Questionnaire (CCSS-NCQ) Revised: Item Response Analysis and Concurrent Validity

Kelly M. Kenzik¹, I-Chan Huang^{1,2}, Tara M. Brinkman^{3,4}, Brandon Baughman⁵, Kirsten K. Ness³, Elizabeth A. Shenkman^{1,2}, Melissa M. Hudson^{3,6}, Leslie L. Robison³, and Kevin R. Krull^{3,4}

¹Department of Health Outcomes and Policy, College of Medicine, University of Florida

²Institute for Child Health Policy, College of Medicine, University of Florida

³Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital

⁴Department of Psychology, St. Jude Children's Research Hospital

⁵Department of Neuropsychology, Semmes Murphey Neurologic & Spine Institute

⁶Department of Oncology, St. Jude Children's Research Hospital

Abstract

Objective—Childhood cancer survivors are at risk for neurocognitive impairment related to cancer diagnosis or treatment. This study refined and further validated the Childhood Cancer Survivor Study Neurocognitive Questionnaire (CCSS-NCQ), a scale developed to screen for impairment in long-term survivors of childhood cancer.

Method—Items related to task efficiency, memory, organization and emotional regulation domains were examined using item response theory (IRT). Data were collected from 833 adult survivors of childhood cancer in the St. Jude Lifetime Cohort Study who completed self-report and direct neurocognitive testing. The revision process included: 1) content validity mapping of items to domains, 2) constructing a revised CCSS-NCQ, 3) selecting items within specific domains using IRT, and 4) evaluating concordance between the revised CCSS-NCQ and direct neurocognitive assessment.

Results—Using content and measurement properties, 32 items were retained (8 items in 4 domains). Items captured low to middle levels of neurocognitive concerns. The latent domain scores demonstrated poor convergent/divergent validity with the direct assessments. Adjusted effect sizes (Cohen's *d*) for agreement between self-reported memory and direct memory assessment were moderate for total recall (ES=0.66), long-term memory (ES=0.63), and short-term memory (ES=0.55). Effect sizes between self-rated task efficiency and direct assessment of attention were moderate for focused attention (ES=0.70) and attention span (ES=0.50), but small

for sustained attention (ES=0.36). Cranial radiation therapy and female gender were associated with lower self-reported neurocognitive function.

Conclusion—The revised CCSS-NCQ demonstrates adequate measurement properties for assessing day-to-day neurocognitive concerns in childhood cancer survivors, and adds useful information to direct assessment.

Improved cancer detection and treatments over the past three decades have resulted in substantial increase of 5-year survival rates for children and adolescents (Howlader et al., 2012). This success has permitted the identification of treatment-related late-effects in long-term survivors (Ness et al., 2008; Oeffinger, Nathan, & Kremer, 2010; Zebrack, Yi, Petersen, & Ganz, 2008). Late-effects include a variety of chronic conditions, including physical, psychological, and social problems (Schultz et al., 2007; Zebrack et al., 2008), as well as neurological impairments (Krull et al., 2008; Packer et al., 2003), all of which have a negative impact on quality of life. Neurocognitive impairments are of particular concern given they occur in over 40% of survivors (Moleski, 2000; R. Mulhern et al., 1998), and are associated with reduced academic and vocational attainment (Gurney et al., 2009; Kirchhoff et al., 2011; Krull et al., 2008; Moore, 2004; Mulhern, Merchant, Gajjar, Reddick, & Kun, 2004).

Risk for neurocognitive impairments is highest among survivors of central nervous system (CNS) tumors (Palmer et al., 2001), those who received CNS radiation (Palmer et al., 2003; Reimers et al., 2003; Ris, Packer, Goldwein, Jones-Wallace, & Boyett, 2001), and those treated with high doses of antimetabolites (Carey et al., 2007; Conklin et al., 2012). Treatment for tumors located in the cerebral hemispheres have been found to be associated with poor IQ, memory, attention, and motor skills (Armstrong et al., 2010; Ater et al., 1996; Robinson et al., 2010). Cranial radiation is associated with brain white matter reduction, particularly in survivors diagnosed at young ages (Armstrong et al., 2013; Mulhern et al., 2004; Palmer et al., 2003). The use of high dose anti-metabolite chemotherapy is also associated with risk for poor neurocognitive outcomes (Anderson & Kunin-Batson, 2009; Edelstein et al., 2011; Reimers et al., 2003).

Neuropsychological testing is the gold standard in evaluation of cognitive late effects of cancer therapy (Krull et al., 2008). However, this testing is time-consuming and difficult to incorporate into routine monitoring in cancer survivors, particularly for cohorts dispersed across wide geographic regions. In addition, this testing is primarily performance-based which does not necessarily represent the patient's perception of daily functioning. In this regard, it is crucial to use self-report measures to capture the patient's perception or neurocognitive concerns and to serve as a tool to screen the patient's neurocognitive deficits. Self-report information of a patient's concerns relating to neurocognitive symptoms may assist in directing further assessment and intervention for those with neurocognitive loss that affects daily function (Atherton & Sloan, 2006).

Rating scales have been developed for assessment of neurocognitive concerns in adults with specific conditions, such as dementia and brain trauma. Examples of such scales include the Cognitive Behavior Rating Scale (CBRS) designed for measuring dementia-like symptoms (Williams, 1987), the Neurobehavioral Functioning Inventory (NFI) for tracking brain-

trauma rehabilitation (Kreutzer, See, & Marwitz, 1999), and the Frontal Systems Behaviors Scale (FrSBe) for assessing frontal lobe damage (Grace & Malloy, 2001). However, these instruments have not demonstrated sensitivity to concerns about functional impairment in childhood cancer survivors, and do not always include assessment of cognitive domains commonly impaired in survivors, such as attention skills or processing speed (Krull et al., 2008). The Childhood Cancer Survivor Study-Neurocognitive Questionnaire (CCSS-NCQ) was developed to assess specific self-reported concerns about cognitive function in long-term survivors of childhood cancer. The CCSS-NCQ was developed in conjunction with the Behavior Rating Inventory of Executive Function – Adult Version (BRIEF-A), using similar items and including novel items specific to outcomes in survivors of childhood cancer (Krull et al., 2008). The CCSS-NCQ has been validated in cohorts of adult survivors of childhood cancer and their siblings, and the instrument score has been found to be impacted by cranial radiation therapy, and related to educational and vocational limitations (Ellenberg et al., 2009; Kadan-Lottick et al., 2010; Kirchhoff et al., 2011).

From a measurement perspective, the aforementioned instruments including the CCSS-NCQ are not without limitations and deserve further investigation. First, the original CCSS-NCQ has not yet been measured against the gold standard direct neurocognitive assessments. It is unclear whether the CCSS-NCQ measures different or similar neurocognitive domains compared to the direct assessments. Second, in the original CCSS-NCQ 19 of the 25 total items were used by the developers to calculate domain scores (Krull et al., 2008). The developers suggest that the initial item selection needs further revision to identify a set of items where all items can be used for scoring to provide maximum information. Finally, the original CCSS-NCQ was developed and validated using classical test theory (CTT). However, CTT retains some limitations because the estimates of item or test properties depend on sample properties and estimates of sample properties depend on item or test properties. With CTT, it is unclear whether a given high score on a single item is truly the result of an individual with fewer neurocognitive concerns, or because of the ability of this item to capture lower levels of neurocognitive concerns (Allen & Yen, 1979/2002). In other words, instrument development based on CTT is often not able to distinguish the self-reported neurocognitive concerns of an individual apart from the capabilities or difficulty of the instruments. As a result, the instrument may be limited in its abilities to describe a broad range of neurocognitive concerns, missing either extreme levels or demonstrating floor/ceiling effects, leading to imprecise outcome measurement (Hambleton, Swaminathan, & Rogers, 1991; Hambleton & Jones, 1993). Revising the CCSS-NCQ using sophisticated measurement methodology and validating the revision against the gold standard neurocognitive testing methods will address the original instrument's limitations and help establish an accurate and clinically relevant measure of neurocognitive concerns.

Item-response theory (IRT) is a modern measurement methodology that features item level analysis and sample independence (Hays, Morales, & Reise, 2000; Reise & Waller, 2009). Each item is characterized by specific item parameters, such as discrimination and difficulty. The item parameters of the instrument and underlying trait of self-reported neurocognitive concerns are calibrated on the same metric, allowing evaluation of an individual's response to specific items given their underlying neurocognitive concerns (Hambleton, 2000; Hays et al., 2000). Importantly, IRT is useful for linking multiple instruments and selecting high

quality items to create an item pool for measuring a broader scope of underlying neurocognitive concerns (Reise & Waller, 2009). In this regard, the use of IRT can improve the measurement capabilities of instruments like the CCSS-NCQ.

The main purpose of this study was to apply IRT methodology to refine and further validate the CCSS-NCQ. We used IRT methodology to help select high quality items, and then validate the revision by comparing it to direct neuropsychological testing. Finally, we examined ratings across clinical and demographic variables associated with concerns about neurocognitive impairment. We hypothesized that this revised CCSS-NCQ would better capture underlying neurocognitive concerns in survivors of childhood cancer.

Methods

Study sample

The survivor sample was taken from the ongoing St. Jude Lifetime Cohort study (SJLIFE), which includes survivors of childhood cancer originally diagnosed and treated at St. Jude Children's Research Hospital (Hudson et al., 2011; Ojha et al., 2013). Enrollment in SJLIFE is limited to those 10 years from diagnosis and 18 years of age. All survivors with a history of cranial radiation therapy, antimetabolite chemotherapy, or neurosurgery undergo comprehensive neuropsychological testing. For the current analyses, we identified 833 survivors (Table 1) who completed the CCSS-NCQ and BRIEF-A and then a direct neurocognitive assessment with a certified examiner. Half of the participants were male (50.7%), and the majority were White (85.1%). Age at time of testing ranged from 18 to 55 years. Approximately 70% were diagnosed with acute lymphoblastic leukemia and 55% received cranial radiation.

Self-report of neurocognitive concerns

The CCSS-NCQ was designed for the Childhood Cancer Survivor Study (CCSS) to assess long-term survivors' neurocognitive skills potentially affected by treatments, including radiation and anti-metabolite chemotherapy (Krull et al., 2008). In its development, some items were included from a pre-published version of the BRIEF-A, in collaboration with the authors, while additional items were newly created to assess processing speed, memory, attention, and academic function. Standardization was conducted in adult survivors of childhood cancer, referenced to age- and sex-matched sibling controls. Based on the exploratory factor analyses, 25 items were retained to measure four domains of neurocognitive concerns: memory, task efficiency, organization, and emotional regulation. Participants are asked to report the degree to which they experience the problem presented by the item over the past 6 months. Response options are on a Likert scale ranging from 1 ("never a problem") to 3 ("often a problem"). Psychometric properties at the domain level were generally acceptable, including indices of construct, concurrent, and discriminative validity.

The BRIEF-A was designed to measure executive functioning. It is a standardized self-report measure validated for adults between the ages of 18 and 90 years of age (Roth, Isquith, & Gioia, 2005). Nine scales (75 items) are included to capture the following specific

domains: inhibit, shift, emotional control, self-monitor, initiate, working memory, plan/organize, task monitor and organization of materials. The nine scales form two higher-order constructs: behavioral regulation (representing inhibit, shift, emotional control, and self-monitor domains) and metacognition (representing initiate, working memory, plan/organize, task monitor, and organization of materials domains). Psychometric properties of this instrument have been previously reported (Roth et al., 2005).

Neuropsychological testing

Trained psychological examiners under the general supervision of a board-certified clinical neuropsychologist conducted direct assessments of neurocognitive functioning. The assessments were designed to measure neurocognitive functioning that may have been affected by treatment modalities or the cancer itself. The clinical assessments included evaluation of: memory, attention, processing speed, and executive functioning. Memory was assessed using the California Verbal Learning Test – Second Edition (CVLT-II) for total recall, short-term free recall, and long-term free recall (Delis, Kramer, Kaplan, & Ober, 2000). Attention and processing speed were assessed using the Trail Making Test Part A (Trails A) (Reitan, 1969), the Conners' Continuous Performance Test – Second Edition (CPT-II; Omissions [OMS] and Variability [VAR] indices) (Conners, 1992), and the Wechsler Adult Intelligence Scale - Third Edition (WAIS-III) Processing Speed Index (Wechsler, 1997). Executive functioning was assessed using the WAIS-III Digit Span Backward subtest, Controlled Oral Word Association test (Verbal Fluency) (Lezak, Howieson, & Loring, 2004), and the Trail Making Test Part B (Trails B) (Reitan, 1969). Age-adjusted standard scores were calculated for the indices described above using normative data from standardization samples. Impairment on indices was defined as a score falling 10th percentile on these age-adjusted scales.

Refining the CCSS-NCQ

The first step in refining the CCSS-NCQ was to establish content validity by mapping items to the common cognitive domains through expert input and panel member meetings. Using the four domains from the CCSS-NCQ as a framework (i.e. task efficiency, memory, organization, and emotional regulation), four panel members with expertise in neuropsychology and psychometrics were asked to review content of the entire set of items (100 items) from the CCSS-NCQ and the BRIEF-A, and assign each item into one of the domains. Any discrepancies in the item assignment between the panel members were addressed and reconciled in group meetings. As a result, 17, 10, 12, and 12 items were assigned to the domains of task efficiency, memory, organization, and emotional regulation, respectively, and used for further psychometric analyses. Forty-nine items were set aside because the panel believed they measured concepts other than the four pre-specified domains.

The second step was to assess the construct of the revised CCSS-NCQ. Dimensionality assessment was conducted to provide a measure of construct validity for a specific domain. For each domain, we used confirmatory factor analysis (CFA) to evaluate unidimensionality and local independence (Hu & Bentler, 1999; Wainer, 1996; Yen, 1993), which are the two major assumptions of IRT modeling. However, item response matrices are rarely strictly

unidimensional in patient-reported outcomes measures such as pain, fatigue, and concerns about neuropsychological functioning (Hays et al., 2000; Reise & Waller, 2009). In this regard, the use of “essential unidimensionality” becomes an acceptable criterion to relax the strict unidimensionality assumption (Strout, 1990). In CFA, comparative Fit Indices (CFI) >0.95 , Tucker-Lewis-Index (TLI) >0.95 , and Root Mean Square Error of Approximation (RMSEA) <0.06 were used to determine the acceptable level of unidimensionality (Hu & Bentler, 1999). Residual correlation <0.2 was used to determine acceptable level of local independency. We deleted items with residual correlations above the acceptable criteria. If the fit indices for the individual domains were suboptimal, additional analyses were conducted to determine whether the essential unidimensionality was held by individual domains, and the degree to which this property biased the estimation of IRT item parameters. First, “problematic” items contributing to the poor fit indices (those with the highest residual correlations and/or highest modification index comparative to the other items) were identified. Second, a bifactor model was constructed within each domain to test whether all items measure an overarching concept (a single general factor) and subsets of items measure multiple sub-concepts (two common factors). The first common factor was measured by the “problematic” items and the second common factor was measured by the remaining items in the domain. If the fit indices of the bifactor model were superior to the fit indices of unidimensional CFA model and the item factor loadings on the general factor of bifactor models were similar relative to the loadings in the unidimensional CFA model, this indicates that the data were essentially unidimensional rather than strictly unidimensional (Reise et al., 2011). If the essential unidimensionality is recognized by the bifactor models, IRT analyses were subsequently conducted to determine whether the 8-item full models were biased by comparing item parameters of the full models to the parameters of the models with each of the problematic items removed (one-by-one from the 8-item model). In each domain, if the item parameters were similar between the 8-item model and the models with each problematic item removed, then the 8-item model was retained. The robust estimations of item parameters will lead to unbiased estimation of TIF/reliability.

For the third step, we used IRT methodology to evaluate measurement properties of each item within individual domains to guide the selection of appropriate items. We used a graded response model (GRM), which is a polytomous IRT model (Samejima, 1996), to accommodate items with more than two response categories. The GRM model estimates the discrimination (a-parameter) and the threshold (or known as difficulty in IRT with two response categories) parameter (b-parameter) for each item. Item discrimination is used to quantify how strongly an item is related to the underlying latent trait of neurocognitive concerns (theta; θ) measured by the CCSS-NCQ (Fayers & Machin, 2007; Hays et al., 2000). Higher discrimination values suggest these items possess greater ability to distinguish subjects with different levels of neurocognitive concerns than items with lower discrimination values (Supplementary Figure 1). The underlying trait represents the subjective neurocognitive concerns such as memory subscale or task efficiency. We treat neurocognitive concerns as an underlying trait because this type of psychological outcome cannot be measured or observed directly. Instead, we use questionnaires to capture the latent concept (i.e., underlying trait). Underlying neurocognitive concerns were estimated by the

GRM along a dimension called theta, and the theta scores were scaled with a mean of 0 and a stand deviation of 1.

Traditionally, for items with dichotomous response categories, the “difficulty” parameter is used to quantify how easy or how difficult it is for an item to measure the intended concept. If items have more than two response categories, item difficulty is typically called “threshold.” For the CCSS-NCQ items which have three response categories, two threshold parameters (i.e., the 1st response category vs. the 2nd category and the 2nd category vs. the 3rd category) are included in GRM. A specific threshold of an item indicates the amount of latent trait at which 50% of subjects choose the designated level of response category or higher. In other words, the threshold parameter represents the level of the underlying neurocognitive concerns necessary for a subject to respond to above a particular category with a probability of 0.50. The values of threshold parameters are calibrated on the same metric as the underlying neurocognitive concerns (Fayers & Machin, 2007; Hays et al., 2000). Location parameter is the mean of threshold parameters of a particular item, which represent the overall difficulty of the item. Given the scores of underlying neurocognitive concerns of our sample fall within the range -3 and +3 units¹ in our study, if the hypothetical item “be able to do calculus in my head” has the highest location at 2.5, this item is considered as a more “difficult” item than other items. Subjects with high latent scores (less or no neurocognitive concerns) would be able to endorse the highest response category.

The concept of threshold and discrimination for each item can be depicted using a category response curve (CRC) which represents the probability of responding to particular category conditional on a certain level of the underlying trait (Embretson & Reise, 2000; Hays et al., 2000) (Supplementary Figure 1). Additionally, item information function (IIF) and test information function (TIF) were estimated to describe the precision of measurement at the item and domain levels, respectively, across different levels of the underlying trait of neurocognitive concerns (Hays et al., 2000; Reise, Ainsworth, & Haviland, 2005).

Additionally, item information function (IIF)² and test information function (TIF)³ were estimated to describe the precision of measurement at the item and domain levels, respectively, across different levels of the underlying trait neurocognitive concerns (Hays et al., 2000; Reise et al., 2005). The TIF is a measure of precision because it has a reverse relationship to the subject's standard error⁴ (Embretson & Reise, 2000).

¹There are no range restrictions on values of underlying cognitive concerns. A range (-3 to +3) was reported to reflect that our estimates fall within this range, and higher scores indicate less neurocognitive concerns.

²IIF for a polytomous IRT model:

$$I_i(\theta) = \sum_{x=0}^m P_{ix}(\theta)^2 \div P_{ix}(\theta) \quad (\text{Embretson \& Reise, 2000})$$

$$I(\theta) = \sum_{i=1}^t I_i(\theta)$$

³TIF (summation of IIFs for a given scale):

Where i indexes the items ($i=1,2,3,\dots,n$); θ represents level of underlying trait; x represents a particular score category; m represents the number of response categories; t represents the number of items; $I_i(\theta)$ is the amount of item i information at an ability level of θ , P_{ix} is the probability of endorsing that response category conditional on the level of the underlying trait.

⁴ $SE(\theta) = 1 \div \left(\sqrt{TIF(\theta)} \right)$

We carefully selected items derived from the two instruments based on the comparisons of location and discrimination parameters, and IIF. If items had similar values in location and discrimination but appeared to have different content the items were retained to ensure adequate domain coverage. If items had similar values in location or discrimination as well as the content, one item was selected after evaluating the range of the latent trait covered by each item using the threshold parameters and IIFs. Ideally, it is desirable to have items covering a larger range of the latent trait (e.g., some items may cover the lower end of the latent trait while other items cover the higher end) and provide the greater level of information. TIF was used to determine the extent to which the specific domains reliably measure different levels of the underlying trait of neurocognitive concerns. In this study, a cut-off value >10.0 was used, which is equivalent to Cronbach's alpha 0.9 in CTT. The relationship between reliability and the TIF is based on the equation

$$0.90 = \left(1 - \left(1 \div \sqrt{10} \right)^2 \right) \text{ (Reeve \& Fayers, 2005).}$$

Validate the revised CCSS-NCQ using clinical neurocognitive assessments

We evaluated the discriminative validity of the revised CCSS-NCQ using a known-groups validation design (Fayers & Machin, 2007) to demonstrate the revised domains' abilities to discriminate between subjects with known levels of neurocognitive status based on direct neuropsychological assessment. In this study, we further evaluated the discriminative validity of the original versus the revised CCSS-NCQ by examining the effect sizes of the original versus the revised CCSS-NCQ for each of the direct neurocognitive assessments. We also evaluated the correlations between the IRT theta estimates and the continuous clinical neurocognitive assessment scores to see whether the domains had divergent/convergent validity.

Multiple linear regression analyses were used to evaluate known-groups validity; each clinical assessment measure was modeled as a dichotomous dependent variable (0 / 1: not impaired / impaired) and the score of the revised CCSS-NCQ domain was modeled as the independent variable. We conducted both unadjusted and adjusted (including current age, sex and race as covariates) regression analyses and reported the differences in neurocognitive concerns measured by the revised CCSS-NCQ domains between subjects with and without neurocognitive impairment as determined by clinical assessment. We also calculated the effect sizes (ES) for known-groups validity, defined as the mean difference in scores of the CCSS-NCQ between subjects with and without impaired neurocognitive functioning for a specific clinical assessment divided by the pooled standard deviation of the scores (Cohen, 1988). Effect sizes <0.2 are classified as negligible, $0.2-0.49$ as small, $0.5-0.79$ as moderate, and >0.8 as large (Cohen, 1988). Effect sizes ≥ 0.5 are considered as a clinically important difference (Norman, Sloan, & Wyrwich, 2003).

Identifying clinical and demographic factors associated with neurocognitive concerns

We examined clinical and demographic factors previously found to be associated with poor neurocognitive outcomes (Askins & Moore, 2008; Ellenberg et al., 2009; Goldsby et al., 2010). We conducted multiple linear regression with each underlying domain score as the dependent variable and the following factors as independent variables: age at diagnosis (0-2

years, 3-5 years, 6-10 years, 11-15 years, and 16-20 years), cranial radiation (yes/ no), IV methotrexate (yes/ no), intrathecal methotrexate (yes/ no), high dose IV methotrexate (yes/ no), relapse of cancer (yes/ no), current age (17-24 years, 25-29, 30-34, 35-40, 40+), sex, and race (White/ Other).

In this study, CFA was conducted using Mplus 7 with the option of the weighted least squares with robust standard errors, mean- and variance- adjusted (WLSMV) to appropriately analyze the categorical data (Muthen & Muthen, 2012), IRT was conducted using Parscale 4.0 (Muraki & Bock, 1997), and other analyses were conducted using SAS Version 9.1 (SAS Institute Inc., Cary, NC).

Results

Refining the CCSS-NCQ

Item mapping resulted in retention of 51 items: 10 items in the memory domain (5 from NCQ and 5 from BRIEF-A), 17 items in the task efficiency domain (9 from NCQ and 8 from BRIEF-A), 12 items in the organization domain (3 from NCQ and 9 from BRIEF-A), and 12 items in the emotional regulation domain (3 from NCQ and 9 from BRIEF-A). For the initial memory 10 item set, CFI and TLI were 0.93 and 0.92, and RMSEA was 0.18. For task efficiency, CFI and TLI were 0.96 and 0.95, and RMSEA was 0.14. For organization, CFI and TLI were 0.95 and 0.94, and RMSEA was 0.16. For emotional regulation, CFI and TLI were 0.97 and 0.96, and RMSEA was 0.10.

Five items were deleted based on residual correlations >0.2 (3 items in task efficiency and 2 items in organization). Residual correlations for the paired items within the same domain were all less than 0.2, suggesting that local independence was acceptable. The fit indices improved slightly for task efficiency (CFI/TLI: 0.95/0.94, RMSEA: 0.10) and organization (CFI/TLI: 0.93/0.91, RMSEA: 0.15), suggesting that the individual domains of the CCSS-NCQ are unidimensional with acceptable measurement errors indicated by CFI and TLI and marginally acceptable errors by RMSEA. The research team would not err on the side of deleting too many items at this stage of CFA because these items were content appropriate. Therefore, these 46 items were retained for IRT analyses.

Based on the item content and measurement properties of item location, thresholds, and discrimination parameters, 14 items were eliminated leaving 32 items (with 8 items for each domain) for the revised CCSS-NCQ. Table 2 shows the measurement properties of the final set of items estimated by the two-parameter IRT model. The fit indices for the 8-item individual domains estimated by the CFA were: CFI/TLI=0.92/0.88 and RMSEA=0.17 for memory; CFI/TLI=0.98/0.97 and RMSEA=0.11 for task efficiency; CFI/TLI=0.92/0.89 and RMSEA=0.17 for organization; CFI/TLI=0.97/0.96 and RMSEA=0.12 for emotional regulation. Additional analyses using the bifactor models suggest that the fit indices were improved in the memory domain (CFI/TLI=1.00/0.99 and RMSEA=0.01), task efficiency domain (CFI/TLI=1.00/0.99 and RMSEA=0.05), organization domain (CFI/TLI=1.00/0.99 and RMSEA=0.02), and emotional regulation domain (CFI/TLI=1.00/0.99 and RMSEA=0.04). In individual domains, the factor loadings of the items on the general factor were similar to the factor loadings in the unidimensional CFA models (e.g., for the

organization domain, the factor loading of BRIEF-7 was 0.71 in unidimensional CFA and 0.73 in the bifactor model). Additional IRT analyses suggest that for each individual domain, regardless of which problematic item was excluded, the discrimination, location and threshold parameters of other items across the 7-item alternate models were very similar (Supplementary Table 2). Within each domain, the item parameters were also highly comparable between the 7-item alternative models and the 8-item full model. These findings suggest that the use of unidimensionality IRT model to fit the data with a property of essential dimensionality still yields a robust estimation of item parameters and TIF/reliability. Figure 1 shows TIF for each domain. The information function across four domains was less than 10. The task efficiency and organization domains had the highest TIF with each domain around 5. However, all TIFs were shifted towards the lower end of the underlying trait, suggesting that the revised CCSS-NCQ provides greater measurement precision (or less measurement error) at low levels of neurocognitive concerns than at high levels.

Validating the revised CCSS-NCQ using clinical neurocognitive assessments

Table 3 shows the discriminative validity (i.e. groups with known impairment based on clinical neurocognitive testing) of the revised CCSS-NCQ. For the memory domain, subjects who were classified as having neurocognitive impairment by clinical assessments demonstrated lower self-reported memory domain scores on the revised CCSS-NCQ compared to those without impairment (all comparisons with effect size >0.20). The associations were highest for the memory domain of the CVLT-II when compared to other clinical assessments, except for the total recall assessment in the self-reported task efficiency domain (0.656 vs. 0.780). Adjusted effect sizes were moderate based on known-groups for total recall (0.656), long-term memory (0.628), and short-memory (0.58) of the CVLT-II, respectively. For the task efficiency domain, subjects who were classified as having neurocognitive impairment by clinical assessments demonstrated lower task efficiency domain scores on the revised CCSS-NCQ compared to those without impairment (effect sizes between 0.358 and 0.780). Adjusted effect sizes were moderate based on known-groups for Trails A and WAIS-III Longest Digit Forward (0.698 and 0.50, respectively), but small for CPT-II OMS and VAR (0.402 and 0.58, respectively). In contrast to our hypothesis that the task efficiency items would strongly discriminate between known-groups determined by the executive functioning measures, effect sizes of the task efficiency domain were moderate based on known-groups of executive functioning measured by the Verbal Fluency, WAIS-III Longest Digit Backward, and Trails B (0.685, 0.5990, and 0.537, respectively).

For the organization domain, subjects who were classified as having neurocognitive impairment by clinical assessments demonstrated lower organization domain scores on the revised CCSS-NCQ compared to those without impairment (all with effect size >0.20 , except long term memory, CPT-II OMS and VAR and Trails B). Adjusted effect sizes were small for memory known-groups measured by the CVLT-II (total recall and short term memory), Trails A, and Verbal Fluency (0.254, 0.287, 0.209, and 0.200, respectively), and negligible with long-term memory, CPT-II OMS and VAR and Trails B.

For the emotional regulation domain, subjects who were classified as having neurocognitive impairment by clinical assessments demonstrated lower emotional regulation domain scores on the revised CCSS-NCQ compared to those without impairment (all with effect size >0.20). Adjusted effect sizes were moderate (0.573) for executive functioning known-group measured by Verbal Fluency and small (0.301 – 0.468) with other clinical assessments.

Table 4 indicates that the memory and task efficiency domains demonstrate the highest correlations with direct assessments (0.26-0.34). The organization domain has negligible correlations (<0.2) with most direct assessments. Supplementary Figure 2 provides a visual representation for the mean latent scores of those impaired and non-impaired based on the direct assessment measure for each of the domains. This figure demonstrates that the discriminative validity of the memory and task efficiency was greater than the emotional regulation and task efficiency domains. Supplementary Table 1 demonstrates that the effect sizes for discriminant validity for the revised CCSS-NCQ against the direct assessment were slightly superior to that of original CCSS-NCQ, especially in the domains of memory and task efficiency.

Factors associated with neurocognitive concerns

Table 5 shows the factors that contributed to neurocognitive concerns measured with the revised CCSS-NCQ. All domains were significantly negatively associated with cranial radiation except for the organization domain, suggesting that those receiving cranial radiation had lower levels on measures of memory, task efficiency, and emotional regulation. Females were more likely to report more neurocognitive concerns compared to males across all domains of the CCSS-NCQ ($p < 0.05$). However, those experiencing a relapse of cancer had significantly less neurocognitive concerns in the domains of memory ($p < 0.05$) and emotional regulation ($p < 0.05$).

Discussion

There are over 300,000 survivors of childhood cancer currently in the US (Howlader et al., 2012), with an estimated 1 in 640 young adults aged 20-39 years being such a survivor (Mariatto et al., 2009). Given the frequent occurrence of neurocognitive impairment in these individuals, a screen instrument is needed to identify those at risk. The original CCSS-NCQ has demonstrated utility in 12 published manuscripts and has contributed to our understanding of late-effects associated with cranial radiation (Armstrong et al., 2010; Kadan-Lottick et al., 2010), psychoactive medication use (Brinkman et al., 2013), sleep and fatigue (Clanton et al., 2011), employment (Kirchhoff et al., 2011), independent living (Kunin-Batson et al., 2011) and marriage (Janson et al., 2009). The current study advances the initial development of the CCSS-NCQ (Krull et al., 2008), analyzing and selecting high quality items from the initial CCSS-NCQ and the BRIEF-A to cover a broader range of scores for the underlying trait of neurocognitive concerns among childhood cancer survivors. Expert panel and IRT methodology were used to generate a 32-item instrument that is comprised of 4 domains of neurocognitive concerns: memory (8 items), task efficiency (8 items), organization (8 items), and emotional regulation (8 items). This revised tool demonstrates acceptable discriminative validity, where greater neurocognitive concerns

in each CCSS-NCQ measured domain were associated with greater neurocognitive impairment measured by the clinical neurocognitive assessments. Female gender and exposure to cranial radiation were two major factors that significantly contributed to more neurocognitive concerns estimated by the refined instrument.

Although the use of IRT methodology to develop and refine patient-reported outcomes instruments for adult cancer patients and survivors is increasing, there are limited studies that apply IRT methodology to childhood survivors (Huang et al., 2011). The major advantage of IRT is the calibration of items from different instruments on the same measurement metric to allow head-to-head comparison of measurement properties at the individual item level. This study provided strong data upon which to conduct item calibration because each subject completed all items of the CCSS-NCQ and the BRIEF-A allowing for precise estimation of individual item parameters.

In a specific domain, the ideal set of items should comprise those that possess higher item discrimination and different item location, and cover a broader range (i.e. threshold values) of underlying trait of neurocognitive concerns. In this study, we found that several items from both instruments were redundant, and that some items from both instruments shared similar content in wording, but possessed different location and/or discrimination. For example, in the organization domain, we retained BRIEF-A 60 (“leave my room or home a mess”) and eliminated BRIEF-A 31 (“lose things, keys, money, wallet”) because, despite the similar content, the discrimination parameter of the BRIEF-A 60 was higher ($a=1.486$) than BRIEF-A 31 ($a=0.945$). As a result, items retained in the revised CCSS-NCQ are expanded in content, and item location and discrimination parameters than the initial CCSS-NCQ. This implies that the item selection process for this revised instrument was based on both empirical evaluation (e.g., IRT analysis) and expert clinical judgment (e.g., panel judgment for the content).

Our results provide a wealth of quality information for each domain of the revised CCSS-NCQ and identify the gaps for future instrument refinement. Quality information include the extent to which the revised instrument was able to measure underlying neurocognitive concerns, distinguish between survivors with high and low levels of neurocognitive concerns, and reliably measure different levels of underlying neurocognitive concerns. Importantly, we found that items in each domain of the original instrument measured only middle levels to high levels of neurocognitive concerns, evidenced by the fact that the location parameter of all items was negative (except NCQ 24). For example, in the memory domain (Table 2), the item from the BRIEF-A 26 (“trouble staying on same topic when talking”) had a lower location parameter $b= -1.337$ for when compared to that of the NCQ 24 (“difficulty solving math problems in my head”) $b=0.403$, which makes clinical sense because solving math problems in one's head seems inherently more difficult than staying on the same topic when talking.

The discrimination parameter for the majority of items was above 1 and this result is acceptable for items to sufficiently differentiate between those with more or less neurocognitive concerns. This magnitude of item discrimination leads to a moderate level of test information function (TIF) curves (or reliability) on each domain. The largest TIF for

each domain was between 4 and 5, and restricted to a narrow range of underlying neurocognitive concerns (e.g., between -1 and 1 for memory domain). This result implies that the use of revised CCSS-NCQ among survivors who have very low levels of neurocognitive concerns or very high levels of concerns may result in measurement error compared to survivors who have moderate levels of neurocognitive concerns..

The revised CCSS-NCQ demonstrates acceptable discriminative validity against gold-standard neuropsychological assessments. The memory domain demonstrates greatest discriminative validity with moderately large effect size for memory assessment by CVLT-II when compared other clinical assessments (except for total recall assessment in the task efficiency domain). Effect sizes for the task efficiency domain were moderately large for memory assessment by CVLT-II, and executive functioning assessment by the Verbal Fluency, WAIS-III Longest Digit Backward, and Trails B, suggesting this scale captures a broader range of function than we hypothesized. The emotional regulation domain demonstrates moderate effect sizes across several neuropsychological assessments despite the fact that there is no way to directly assess emotional functioning with a clinical instrument. The highest effect size was found within the executive functioning domain. Our finding also suggests there is a moderate association between emotional regulation and attention and memory domains of clinical assessment, consistent with previous research (Christopher & MacDonald, 2005; DeLuca et al., 2005). The moderate effect sizes may be partially explained by the different cancer diagnoses and treatment methods that characterize our sample and caution should be taken when comparing findings across studies.

The convergent/divergent validity of the revised CCSS-NCQ was less satisfactory compared to the discriminative validity analysis. However, for both the convergent/divergent validity and discriminative validity analyses, the highest values (i.e., correlations and effect sizes, respectively) were found in the memory and task efficiency domains. The weak magnitude of the correlations between the latent domain scores and the continuous scores of the direct neurocognitive assessments indicates that the self-reported neurocognitive concerns might be measuring day-to-day functional concerns of long-term childhood cancer survivors which are not necessarily captured by the clinical assessments. Our findings regarding an association between greater neurocognitive concerns and cranial radiation and female gender are in line with previous studies that measured neuropsychological functioning based on clinical neurocognitive performance assessment (Reimers et al., 2003; Smibert, Anderson, Godber, & Ekert, 1996; Spiegler, Bouffet, Greenberg, Rutka, & Mabbott, 2004) and female gender (Barkley, 1997; Goldsby et al., 2010; Mulhern et al., 2004). These studies reported that impaired neurocognitive functioning of childhood cancer survivors is associated with cranial radiation therapy and female gender. Our finding that age at diagnosis was not significantly associated with neurocognitive concerns was consistent with the study by Ellenberg and colleagues which also examined neurocognitive concerns measured by the original CCSS-NCQ (Ellenberg et al., 2009). However, this non-significant finding was in contrast to the significant relationship between age at diagnosis and neurocognitive status reported in other studies that utilized clinical neurocognitive performance scales to determine neurocognitive status (Askins & Moore, 2008; Mulhern et al., 2001; Sands et al., 2001).

The surprising association between relapse of cancer and lower concerns for neurocognitive functioning in the memory and emotional regulation domains is in line with one review study suggesting that adult patients with recurrent brain tumors (gliomas) demonstrated stable quality of life and improved neurocognitive function with the receipt of one of the new tumor treatment methods tumor (Henriksson, Asklund, & Poulsen, 2011). In the childhood cancer survivor population, it is possible that new treatment methods received during the relapse of cancer may have been better targeted to prevent neurocognitive decline, and may be perceived by patients as being less toxic. Patients who have survived relapse may also come to perceive themselves as more resilient than survivors of a single cancer event. This issue has not been thoroughly examined and more studies are needed to understand the mechanisms through which relapse of cancer and related treatment influence neurocognitive functioning among childhood cancer survivors.

The findings of this study have several implications for both the previously published studies using the CCSS-NCQ and for future research where neurocognitive concerns are measured among childhood cancer survivors. First, the use of qualitative (item mapping) and quantitative (IRT) approaches improves the quality of the original CCSS-NCQ since the discriminative validity related to clinical assessment measures for the revised CCSS-NCQ was improved compared to the original CCSS-NCQ, particularly in memory and task efficiency. The overall magnitude of the effect sizes in both the original and revised CCSS-NCQ were small to moderate, suggesting that the instrument may be measuring different concepts of neurocognitive concerns that are not measured by the clinical assessments.

Second, some of the measurement properties of the revised CCSS-NCQ are still not satisfied (e.g., TIF and dimensionality assessment) or not evaluated in this stage (e.g., responsiveness to change). The final CFA reveals one fit index (i.e., the RMSEA) is slightly higher than the desired value. It is commonly acknowledged that the assumption of unidimensionality for patient-reported outcomes (PROs) cannot be strictly met because PRO items more or less capture multiple tasks (e.g., “lose things, keys, money, and wallet” may capture not only memory issues but also other hidden psycho-behavioral issues) and several factors will affect the test performance including cognitive, psychological, and test-taking factors (Hambleton et al., 1991). A previous study has called for a careful use of CFA fit indices to evaluate dimensionality of PRO data because these indices were sensitive to the distribution of data and the number of items (Cook, Kallen, & Amtmann, 2009). Given minor deviations in the strict assumptions of IRT often occur when items measures complex PRO (Buysse et al., 2010; Reise, Waller, & Comrey, 2000; Revicki et al., 2009), we conducted bifactor modeling to evaluate whether the IRT parameters will be unbiased in the presence of these deviations. The bifactor modeling is useful to identify whether the remaining variance reflected in the suboptimal fit indices is due to the fact that some items not only measure the overarching concept, but also measure another concept that was caused by multiple items potentially tapping into similar aspects of the latent concept (Cook et al., 2009; Reise et al., 2011). We also specifically investigated whether the violation of strict unidimensionality will bias the IRT item parameters by removing problematic items one-by-one. Theoretical studies have suggested a serious violation of unidimensionality can be evident if there is a substantial difference in the item discrimination parameters derived from the models that include locally dependent (or content overlap) items and the model on which the dependent

items are removed (Chen & Thissen, 1997). In this study, we found that the item parameters of a specific domain are robust between the 8-item full model and the 7-item alternative models (one-by-one exclusion of problematic items). This finding echoes previous simulation studies suggesting that the use of a unidimensional IRT models is robust to moderate degree of multidimensional data generated by CFA (Harrison, 1986; Kirisci, Hsu, & Yu, 2001). It is not surprising that the items in the revised CCSS-NCQ were not strictly unidimensional because these items were generated from homologous tools (the original CCSS-NCQ and the BRIEF-A) and the item contents were homogeneous. Although our expert team has prudently removed items with overlapping content, we still need input from survivors to understand their cognitive response to and interpretation of the item contents of the revised CCSS-NCQ. It is also important to develop and validate new, distinct items alongside the extent items to strengthen the unidimensionality of individual domains.

Although this study provides a solid foundation to measure neurocognitive concerns for childhood cancer survivors, we should consider several study limitations. First, our sample is comprised of homogenous race/ ethnicity (85.1% White) and is restricted to long-term childhood cancer survivors enrolled in the St. Jude Lifetime Cohort. Therefore, the generalizability of our finding may be limited to other groups of childhood cancer survivors. However, this study does include individuals who have received more recent treatment agents, such as high-dose methotrexate, which provides a strong foundation to validate the CCSS-NCQ. Second, the selection of items merely involves input of content experts and IRT methodology. We did not explicitly use qualitative methods to receive insight from survivors for investigating their understanding of item meaning and/ or generating new items. Third, we are unable to evaluate the sensitivity of change in neurocognitive concerns over time given the cross-sectional study design. In conclusion, this study extends the previous version of the CCSS-NCQ (Krull et al., 2008) by using expert opinion and rigorous IRT methodology. We demonstrate that the use of different items from the initial CCSS-NCQ and the BRIEF-A can help improve measurement properties compared to the use of either instrument alone. Future studies should collect data of self-report neurocognitive concerns together with and clinical assessment over time to provide evidence of longitudinal validity and sensitivity to change associated with interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Allen, M.; Yen, W. Introduction to measurement theory. Long Grove, IL: Waveland Press; 1979/2002.
- Anderson FS, Kunin-Batson AS. Neurocognitive late effects of chemotherapy in children: The past 10 years of research on brain structure and function. *Pediatric Blood & Cancer*. 2009; 52(2):159–164.10.1002/pbc.21700 [PubMed: 18680151]
- Armstrong GT, Reddick WE, Petersen RC, Santucci A, Zhang N, Srivastava D, Krull KR. Evaluation of memory impairment in aging adult survivors of childhood acute lymphoblastic leukemia treated with cranial radiotherapy. *Journal of the National Cancer Institute*. 2013; 105(12):899–907.10.1093/jnci/djt089 [PubMed: 23584394]

- Armstrong GT, Jain N, Liu W, Merchant TE, Stovall M, Srivastava DK, Krull KR. Region-specific radiotherapy and neuropsychological outcomes in adult survivors of childhood CNS malignancies. *Neuro-Oncology*. 2010; 12(11):1173–1186.10.1093/neuonc/noq104 [PubMed: 20716593]
- Askins MA, Moore BD 3rd. Preventing neurocognitive late effects in childhood cancer survivors. *Journal of Child Neurology*. 2008; 23(10):1160–1171.10.1177/0883073808321065 [PubMed: 18952582]
- Ater JL, Moore BD 3rd, Francis DJ, Castillo R, Slopis J, Copeland DR. Correlation of medical and neurosurgical events with neuropsychological status in children at diagnosis of astrocytoma: Utilization of a neurological severity score. *Journal of Child Neurology*. 1996; 11(6):462–469. Retrieved from <http://jcn.sagepub.com/>. [PubMed: 9120225]
- Atherton PJ, Sloan JA. Rising importance of patient-reported outcomes. *The Lancet Oncology*. 2006; 7(11):883–884.10.1016/S1470-2045(06)70914-7 [PubMed: 17081911]
- Barkley RA. Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*. 1997; 121(1):65–94.10.1037/0033-2909.121.1.65 [PubMed: 9000892]
- Brinkman TM, Zhang N, Ullrich NJ, Brouwers P, Green DM, Srivastava DK, Krull KR. Psychoactive medication use and neurocognitive function in adult survivors of childhood cancer: A report from the childhood cancer survivor study. *Pediatric Blood & Cancer*. 2013; 60(3):486–493.10.1002/pcb.24255 [PubMed: 22848025]
- Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, Pilkonis PA. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep*. 2010; 33(6):781–792. Retrieved from <http://www.journalsleep.org/>. [PubMed: 20550019]
- Carey ME, Hockenberry MJ, Moore IM, Hutter JJ, Krull KR, Pasvogel A, Kaemingk KL. Brief report: Effect of intravenous methotrexate dose and infusion rate on neuropsychological function one year after diagnosis of acute lymphoblastic leukemia. *Journal of Pediatric Psychology*. 2007; 32(2): 189–193.10.1093/jpepsy/jsj114 [PubMed: 16675716]
- Chen W, Thissen D. Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*. 1997; 22(3):265–289. Retrieved from <http://www.jstor.org/stable/1165285>.
- Christopher G, MacDonald J. The impact of clinical depression on working memory. *Cognitive Neuropsychiatry*. 2005; 10(5):379–399.10.1080/13546800444000128 [PubMed: 16571468]
- Clanton NR, Klosky JL, Li C, Jain N, Srivastava DK, Mulrooney D, Krull KR. Fatigue, vitality, sleep, and neurocognitive functioning in adult survivors of childhood cancer. *Cancer*. 2011; 117(11): 2559–2568.10.1002/cncr.25797 [PubMed: 21484777]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: L. Erlbaum Associates; 1988.
- Conklin HM, Krull KR, Reddick WE, Pei D, Cheng C, Pui CH. Cognitive outcomes following contemporary treatment without cranial irradiation for childhood acute lymphoblastic leukemia. *Journal of the National Cancer Institute*. 2012; 104(18):1386–1395.10.1093/jnci/djs344 [PubMed: 22927505]
- Conners, CK. *Conners' Continuous Performance Test*. Toronto, Canada: Multi-Health Systems; 1992.
- Cook KF, Kallen MA, Amtmann D. Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research*. 2009; 18(4):447–460.10.1007/s11136-009-9464-4 [PubMed: 19294529]
- Delis, DC.; Kramer, JK.; Kaplan, E.; Ober, BA. *California Verbal Learning Test®*. Second. San Antonio, TX: The Psychological Corporation; 2000. CVLT®-II
- DeLuca AK, Lenze EJ, Mulsant BH, Butters MA, Karp JF, Dew MA, Reynolds CF 3rd. Comorbid anxiety disorder in late life depression: Association with memory decline over four years. *International Journal of Geriatric Psychiatry*. 2005; 20(9):848–854.10.1002/gps.1366 [PubMed: 16116585]
- Edelstein K, D'agostino N, Bernstein LJ, Nathan PC, Greenberg ML, Hodgson DC, Spiegler BJ. Long-term neurocognitive outcomes in young adult survivors of childhood acute lymphoblastic

- leukemia. *Journal of Pediatric Hematology/Oncology*. 2011; 33(6):450–458.10.1097/MPH.0b013e31820d86f2 [PubMed: 21646917]
- Ellenberg L, Liu Q, Gioia G, Yasui Y, Packer RJ, Mertens A, Zeltzer LK. Neurocognitive status in long-term survivors of childhood CNS malignancies: A report from the childhood cancer survivor study. *Neuropsychology*. 2009; 23(6):705–717.10.1037/a0016674 [PubMed: 19899829]
- Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Mahwah, N.J.: L. Erlbaum Associates; 2000. Retrieved from <http://www.netLibrary.com/urlapi.asp?action=summary&v=1&bookid=44641>
- Fayers, PM.; Machin, D. *Quality of life: The assessment, analysis, and interpretation of patient-reported outcomes*(Second ed). Chichester; Hoboken, NJ: John Wiley & Sons; 2007.
- Goldsby RE, Liu Q, Nathan PC, Bowers DC, Yeaton-Massey A, Raber SH, Packer RJ. Late-occurring neurologic sequelae in adult survivors of childhood acute lymphoblastic leukemia: A report from the childhood cancer survivor study. *Journal of Clinical Oncology*. 2010; 28(2):324–331.10.1200/JCO.2009.22.5060 [PubMed: 19917844]
- Grace, J.; Malloy, PF. *Frontal Systems Behavior Scale*. Lutz, FL: Psychological Assessment Resources, Inc; 2001.
- Gurney JG, Krull KR, Kadan-Lottick N, Nicholson HS, Nathan PC, Zebrack B, Ness KK. Social outcomes in the childhood cancer survivor study cohort. *Journal of Clinical Oncology*. 2009; 27(14):2390–2395.10.1200/JCO.2008.21.1458 [PubMed: 19224833]
- Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practices*. 1993; 12:38–47.10.1111/j.1745-3992.1993.tb00543.x
- Hambleton, RK.; Swaminathan, H.; Rogers, H. *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications; 1991.
- Hambleton RK. Emergence of item response modeling in instrument development and data analysis. *Medical Care*. 2000; 38(9 Suppl):II60–5. Retrieved from <http://www.jstor.org/stable/3768063>. [PubMed: 10982090]
- Harrison DA. Robustness of IRT parameter estimation to violations of the unidimensionality assumption. *Journal of Educational Statistics*. 1986; 11(2):91–115. Retrieved from <http://www.jstor.org/stable/1164972>.
- Hattie J. Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*. 1985; 9(2):139–164.10.1177/014662168500900204
- Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Medical Care*. 2000; 38(9 Suppl):II28–42. [PubMed: 10982088]
- Henriksson R, Asklund T, Poulsen HS. Impact of therapy on quality of life, neurocognitive function and their correlates in glioblastoma multiforme: A review. *Journal of Neuro-Oncology*. 2011; 104(3):639–646.10.1007/s11060-011-0565-x [PubMed: 21468776]
- Howlader, N.; Noone, AM.; Krapcho, M.; Neyman, N.; Aminou, R.; Altekruse, SF.; Cronin, KA. *SEER cancer statistics review, 1975-2009 (vintage 2009 populations)*. Bethesda, MD: National Cancer Insititute; 2012.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6(1):1–55.10.1080/10705519909540118
- Huang IC, Quinn GP, Wen PS, Shenkman EA, Revicki DA, Krull K, Shearer PD. Using three legacy measures to develop a health-related quality of life tool for young adult survivors of childhood cancer. *Quality of Life Research*. 2011.10.1007/s11136-011-0055-9
- Janson C, Leisenring W, Cox C, Termuhlen AM, Mertens AC, Whitton JA, Kadan-Lottick NS. Predictors of marriage and divorce in adult survivors of childhood cancers: A report from the childhood cancer survivor study. *Cancer Epidemiology Biomarkers & Prevention*. 2009; 18(10):2626–2635.10.1158/1055-9965.EPI-08-0959
- Kadan-Lottick NS, Zeltzer LK, Liu Q, Yasui Y, Ellenberg L, Gioia G, Krull KR. Neurocognitive functioning in adult survivors of childhood non-central nervous system cancers. *Journal of the National Cancer Institute*. 2010; 102(12):881–893.10.1093/jnci/djq156 [PubMed: 20458059]

- Kirchhoff AC, Krull KR, Ness KK, Armstrong GT, Park ER, Stovall M, Leisenring W. Physical, mental, and neurocognitive status and employment outcomes in the childhood cancer survivor study cohort. *Cancer Epidemiology, Biomarkers & Prevention*. 2011; 20(9):1838–1849.10.1158/1055-9965.EPI-11-0239
- Kirisci L, Hsu T, Yu L. Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*. 2001; 25(2):146–162.10.1177/01466210122031975
- Kreutzer, JS.; See, RT.; Marwitz, JH. Neurobehavioral functioning inventory. San Antonio, TX: The Psychological Corporation; 1999.
- Krull KR, Gioia G, Ness KK, Ellenberg L, Recklitis C, Leisenring W, Zeltzer L. Reliability and validity of the childhood cancer survivor study neurocognitive questionnaire. *Cancer*. 2008; 113(8):2188–2197.10.1002/cncr.23809 [PubMed: 18792068]
- Kunin-Batson A, Kadan-Lottick N, Zhu L, Cox C, Bordes-Edgar V, Srivastava DK, Krull KR. Predictors of independent living status in adult survivors of childhood cancer: A report from the childhood cancer survivor study. *Pediatric Blood & Cancer*. 2011; 57(7):1197–1203.10.1002/pbc.22982 [PubMed: 21294244]
- Lezak, MD.; Howieson, DB.; Loring, DW. Neuropsychological assessment. 4th. New York, NY: Oxford University Press; 2004.
- Mariotto AB, Rowland JH, Yabroff KR, Scoppa S, Hachey M, Ries L, Feuer EJ. Long-term survivors of childhood cancers in the United States. *Cancer Epidemiology, Biomarkers & Prevention*. 2009; 18(4):1033–1040.10.1158/1055-9965.EPI-08-0988
- Moleski M. Neuropsychological, neuroanatomical, and neurophysiological consequences of CNS chemotherapy for acute lymphoblastic leukemia. *Archives of Clinical Neuropsychology*. 2000; 15(7):603–630.10.1016/S0887-6177(99)00050-5 [PubMed: 14590198]
- Moore BD. Neurocognitive outcomes in survivors of childhood cancer. *Journal of Pediatric Psychology*. 2004; 30(1):51–63.10.1093/jpepsy/jsi016 [PubMed: 15610985]
- Mulhern RK, Merchant TE, Gajjar A, Reddick WE, Kun LE. Late neurocognitive sequelae in survivors of brain tumours in childhood. *The Lancet Oncology*. 2004; 5(7):399–408.10.1016/S1470-2045(04)01507-4 [PubMed: 15231246]
- Mulhern RK, Palmer SL, Reddick WE, Glass JO, Kun LE, Taylor J, Gajjar A. Risks of young age for selected neurocognitive deficits in medulloblastoma are associated with white matter loss. *Journal of Clinical Oncology*. 2001; 19(2):472–479. Retrieved from <http://jco.ascopubs.org/content/19/2/472.long>. [PubMed: 11208841]
- Mulhern R, Kepner J, Thomas P, Armstrong F, Friedman H, Kun L. Neuropsychologic functioning of survivors of childhood medulloblastoma randomized to receive conventional or reduced-dose craniospinal irradiation: A pediatric oncology group study. *Journal of Clinical Oncology*. 1998; 16(5):1723–1728. Retrieved from <http://jco.ascopubs.org/content/16/5/1723.long>. [PubMed: 9586884]
- Muraki, E.; Bock, RD. PARSCALE: IRT item analysis and test scoring for rating scale data. Chicago: Scientific Software; 1997.
- Muthen, B.; Muthen, LK. Mplus version 7.1. Los Angeles, CA: Muthen & Muthen; 2012.
- Ness KK, Gurney JG, Zeltzer LK, Leisenring W, Mulrooney DA, Nathan PC, Mertens AC. The impact of limitations in physical, executive, and emotional function on health-related quality of life among adult survivors of childhood cancer: A report from the childhood cancer survivor study. *Archives of Physical Medicine and Rehabilitation*. 2008; 89(1):128–136.10.1016/j.apmr.2007.08.123 [PubMed: 18164342]
- Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care*. 2003; 41(5):582–592.10.1097/01.MLR.0000062554.74615.4C [PubMed: 12719681]
- Oeffinger KC, Nathan PC, Kremer LC. Challenges after curative treatment for childhood cancer and long-term follow up of survivors. *Hematology/oncology Clinics of North America*. 2010; 24(1): 129–149.10.1016/j.hoc.2009.11.013 [PubMed: 20113899]
- Packer RJ, Gurney JG, Punyko JA, Donaldson SS, Inskip PD, Stovall M, Robison LL. Long-term neurologic and neurosensory sequelae in adult survivors of a childhood brain tumor: Childhood

- cancer survivor study. *Journal of Clinical Oncology*. 2003; 21(17):3255–3261.10.1200/JCO.2003.01.202 [PubMed: 12947060]
- Palmer SL, Gajjar A, Reddick WE, Glass JO, Kun LE, Wu S, Mulhern RK. Predicting intellectual outcome among children treated with 35-40 Gy craniospinal irradiation for medulloblastoma. *Neuropsychology*. 2003; 17(4):548–555.10.1037/0894-4105.17.4.548 [PubMed: 14599268]
- Palmer SL, Goloubeva O, Reddick WE, Glass JO, Gajjar A, Kun L, Mulhern RK. Patterns of intellectual development among survivors of pediatric medulloblastoma: A longitudinal analysis. *Journal of Clinical Oncology*. 2001; 19(8):2302–2308. Retrieved from <http://jco.ascopubs.org/content/19/8/2302.long>. [PubMed: 11304784]
- Reeve, BB.; Fayers, P. Chapter 1 applying item response theory modelling for evaluating questionnaire item and scale properties. In: Fayers, P.; Hays, R., editors. *Assessing quality of life in clinical trials: Methods and practice*. New York: Oxford University Press; 2005. p. 55-73.
- Reimers TS, Ehrenfels S, Mortensen EL, Schmiegelow M, Sonderkaer S, Carstensen H, Muller J. Cognitive deficits in long-term survivors of childhood brain tumors: Identification of predictive factors. *Medical and Pediatric Oncology*. 2003; 40(1):26–34.10.1002/mpo.10211 [PubMed: 12426683]
- Reise SP, Ventura J, Keefe RS, Baade LE, Gold JM, Green MF, Bilder R. Bifactor and item response theory analyses of interviewer report scales of cognitive impairment in schizophrenia. *Psychological Assessment*. 2011; 23(1):245–261. 10.1037/a0021501. 10.1037/a0021501 [PubMed: 21381848]
- Reise SP, Waller NG. Item response theory and clinical measurement. *Annual Review of Clinical Psychology*. 2009; 5:27–48.10.1146/annurev.clinpsy.032408.153553
- Reise SP, Ainsworth AT, Haviland MG. Item response theory. *Current Directions in Psychological Science*. 2005; 14(2):95–101.10.1111/j.0963-7214.2005.00342.x
- Reise SP, Waller NG, Comrey AL. Factor analysis and scale revision. *Psychological Assessment*. 2000; 12(3):287–297.10.1037/1040-3590.12.3.287 [PubMed: 11021152]
- Reitan, R. *Manual for administration of neuropsychological test batteries on adults and children*. Bloomington: Indiana University Press; 1969.
- Revicki DA, Chen W, Harnam N, Cook KF, Amtmann D, Callahan LF, Keefe FJ. Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain*. 2009; 146(1-2):158–169.10.1016/j.pain.2009.07.029 [PubMed: 19683873]
- Ris MD, Packer R, Goldwein J, Jones-Wallace D, Boyett JM. Intellectual outcome after reduced-dose radiation therapy plus adjuvant chemotherapy for medulloblastoma: A children's cancer group study. *Journal of Clinical Oncology*. 2001; 19(15):3470–3476. Retrieved from <http://jco.ascopubs.org/content/19/15/3470.long>. [PubMed: 11481352]
- Robinson KE, Livesay KL, Campbell LK, Scaduto M, Cannistraci CJ, Anderson AW, Compas BE. Working memory in survivors of childhood acute lymphocytic leukemia: Functional neuroimaging analyses. *Pediatric Blood & Cancer*. 2010; 54(4):585–590.10.1002/pbc.22362 [PubMed: 19953649]
- Roth, RM.; Isquith, PK.; Gioia, GA. *Behavior rating inventory of executive functioning - adult version*. Lutz, FL: Psychological Assessment Resources, Inc; 2005.
- Samejima, F. The graded response model. In: van der Linden, WJ.; Hambleton, RK., editors. *Handbook of modern item response theory*. New York, NY: Springer; 1996. p. 85-100.
- Sands SA, Kellie SJ, Davidow AL, Diez B, Villablanca J, Weiner HL, Finlay JL. Long-term quality of life and neuropsychologic functioning for patients with CNS germ-cell tumors: From the first international CNS germ-cell tumor study. *Neuro-Oncology*. 2001; 3(3):174–183.10.1093/neuonc/3.3.174 [PubMed: 11465398]
- Schultz KA, Ness KK, Whitton J, Recklitis C, Zebrack B, Robison LL, Mertens AC. Behavioral and social outcomes in adolescent survivors of childhood cancer: A report from the childhood cancer survivor study. *Journal of Clinical Oncology*. 2007; 25(24):3649–3656.10.1200/JCO.2006.09.2486 [PubMed: 17704415]
- Smibert E, Anderson V, Godber T, Ekert H. Risk factors for intellectual and educational sequelae of cranial irradiation in childhood acute lymphoblastic leukaemia. *British Journal of Cancer*. 1996; 73(6):825–830.10.1038/bjc.1996.145 [PubMed: 8611389]

- Spiegler BJ, Bouffet E, Greenberg ML, Rutka JT, Mabbott DJ. Change in neurocognitive functioning after treatment with cranial radiation in childhood. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*. 2004; 22(4):706–713.10.1200/JCO.2004.05.186 [PubMed: 14966095]
- Strout W. A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. *Psychometrika*. 1990; 55:293–326.10.1007/BF02295289
- Wainer HH. How is reliability related to the quality of test scores? what is the effect of local dependence on reliability? *Educational Measurement, Issues and Practice*. 1996; 15(1):22–29.10.1111/j.1745-3992.1996.tb00803.x
- Weschler, D. Wechsler intelligence scale for children (WAIS-III). Third. New York, NY: The Psychological Corporation; 1997.
- Williams, JM. Cognitive behavior rating scales research edition. Lutz, FL: Psychological Assessment Resources, Inc; 1987.
- Yen WM. Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*. 1993; 30(3):187–213. Retrieved from <http://www.jstor.org/stable/1435043>.
- Zebrack BJ, Yi J, Petersen L, Ganz PA. The impact of cancer and quality of life for long-term survivors. *Psycho-Oncology*. 2008; 17(9):891–900.10.1002/pon.1300 [PubMed: 18050153]

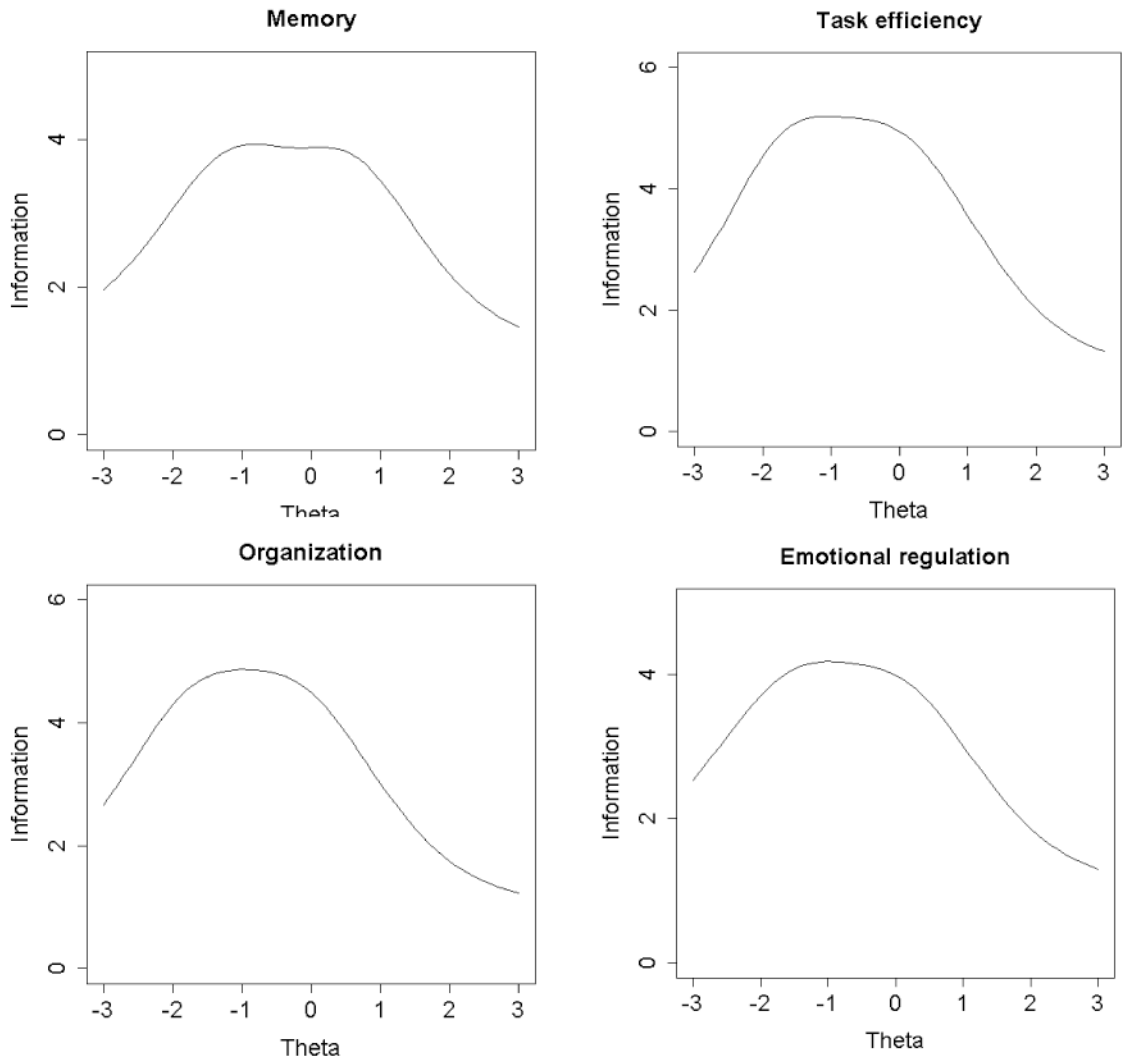


Figure 1. Test information function (TIF) of revised the CCSS-NCQ domains. Each TIF is made up of the sum of item information function (IIF) and represents the total amount of information provided by the domains. The amount information provided by each item is determined by the size of the discrimination or slope (i.e., height of curve) and the spread of the category

thresholds.
$$(IIF: I_i(\theta) = \sum_{x=0}^m P_{ix}(\theta)^2 \div P_{ix}(\theta); TIF: I(\theta) = \sum_{i=t} I_i(\theta) \text{ (Embretson \& Reise, 2000)})$$

Table 1
Sample characteristics (n=833)

Variables	Frequency (%)
Current age (years)	
17-24	124 (14.9)
25-29	178 (21.3)
30-34	199 (23.9)
35-40	174 (20.9)
40+	158 (18.9)
Gender (male)	422 (50.7)
Race	
White non-Hispanic	710 (85.1)
Black non-Hispanic	77 (9.2)
Hispanic	19 (2.3)
American Indian or Alaskan Native	8 (1.0)
Asian	6 (0.7)
Pacific Islander	2 (0.2)
Other	12 (1.4)
Age at diagnosis (years)	
0-2	95 (11.4)
3-5	294 (32.3)
6-10	214 (25.7)
11-15	160 (19.2)
16-20	71 (8.5)
Primary diagnosis	
Acute Lymphoblastic Leukemia	575 (69.0)
CNS malignancy	92 (11.0)
Other malignancy ^a	166 (19.9)
Cranial radiation	
Yes	460 (55.2)
Received intravenous methotrexate (MTX)	
Yes	297 (35.7)
Received intrathecal methotrexate (MTX)	
Yes	593 (71.2)
Received high dose MTX	
Yes	365 (43.9)
Relapse	
Yes	160 (19.2)

Note:

^aOther malignancy includes retinoblastoma (n=52), osteosarcoma (n=40), acute myeloid leukemia (n=24), non-Hodgkin lymphoma (n=22), rhabdomyosarcoma (n=10), Ewing's family of tumors (n=5), Hodgkin's disease (n=3), and other (n=10)

Table 2

Item characteristics of the revised CCSS-NCQ

Item	Contents	d^a	Threshold 1 ^b	b^c	Threshold 2 ^b
Memory					
BRIEF-A 26	Trouble staying on the same topic when talking	0.676	-2.293	-1.337	-0.381
BRIEF-A 55	People say I am easily distracted	0.753	-1.878	-1.074	-0.270
NCQ 13	Forget what I am doing in the middle of things	1.255	-1.660	-0.700	0.260
BRIEF-A 35	Short attention span	0.691	-1.692	-0.652	0.388
NCQ 5	Forgets instructions easily	1.599	-1.373	-0.503	0.367
NCQ 20	Trouble remembering things even for a few minutes	1.750	-0.934	-0.154	0.626
NCQ 7	Difficulty recalling things I learned before	1.372	-1.087	-0.034	1.019
NCQ 24	Difficulty solving math problems in my head	0.669	-0.626	0.403	1.432
Task efficiency					
NCQ 14	Problems with self-motivation	1.174	-2.107	-1.232	-0.357
NCQ 6	Problems completing my work	1.627	-2.043	-1.187	-0.331
NCQ 21	Trouble prioritizing	1.284	-1.913	-1.041	-0.169
NCQ 16	Easily overwhelmed	1.101	-1.699	-0.784	0.131
NCQ 17	Trouble multi-tasking	1.397	-1.492	-0.726	0.040
NCQ 23	Slower than others completing work	1.494	-1.347	-0.608	0.131
NCQ 2	Takes longer to complete work	1.620	-1.294	-0.473	0.348
NCQ 25	Doesn't work well under pressure	1.310	-1.023	-0.296	0.431
Organization					
BRIEF-A 40	Leave the bathroom a mess	0.846	-2.498	-1.494	-0.490
BRIEF-A 71	Trouble organizing work	1.277	-2.048	-1.227	-0.406
BRIEF-A 66	Problems organizing activity	1.093	-1.907	-1.122	-0.337
NCQ 19	Desk/workspace a mess	1.315	-1.756	-1.003	-0.250
NCQ 12	Trouble finding things in bedroom	0.939	-1.850	-0.910	0.030
NCQ 4	Disorganized	1.359	-1.533	-0.648	0.237
BRIEF-A 60	Leave my room or home a mess	1.534	-1.303	-0.459	0.385
BRIEF-A 7	Messy closet	1.027	-1.009	-0.288	0.433
Emotional regulation					

Item	Contents	a^a	Threshold 1 b	b^c	Threshold 2 b
BRIEF-A 19	Emotional outburst for little reason	1.691	-1.737	-1.096	-0.455
BRIEF-A 1	Have angry outbursts	0.995	-2.207	-1.01	0.187
BRIEF-A 33	Overreact to small problems	1.552	-1.686	-0.868	-0.050
BRIEF-A 28	React more emotionally to situations than my friends	1.244	-1.575	-0.823	-0.071
NCQ 9	Moods changes frequently	1.196	-1.627	-0.805	0.017
BRIEF-A 12	Overreact emotionally	1.505	-1.418	-0.593	0.232
NCQ 1	Gets upset easily	1.377	-1.480	-0.474	0.532
NCQ 8	Gets frustrated easily	1.335	-1.244	-0.290	0.664

Note:

^aThe discrimination (a) parameter describes how strongly an item is related to the latent trait.

^bThe threshold values indicate the extent of the latent trait covered by the item.

^cThe location (b) parameter describes how easy or difficult it is for a subject to endorse an item.

Table 3

Discriminative validity of the revised CCSS-NCQ

	Impaired ^a	Not impaired ^d	Difference ^b (effect size) ^d	t	Difference ^c (effect size)	t
Memory						
Memory assessment measures						
Total recall	-0.473 (0.938)	0.099 (0.885)	0.572 ^{***} (0.639)	7.28	0.570 ^{***} (0.656)	7.45
Short-term memory	-0.432 (0.932)	0.077 (0.897)	0.508 ^{***} (0.562)	6.14	0.482 ^{***} (0.548)	5.95
Long-term memory	-0.442 (0.930)	0.124 (0.879)	0.565 ^{***} (0.634) ^{**}	7.80	0.545 ^{***} (0.628)	7.67
Attention & processing speed assessment measures						
Trails A	-0.248 (0.961)	0.040 (0.905)	0.289 ^{***} (0.316)	3.62	0.338 ^{***} (0.380)	4.32
CPT-II OMS	-0.247 (0.966)	0.032 (0.909)	0.279 ^{**} (0.304)	3.09	0.215 [*] (0.240)	2.42
CPT-II VAR	-0.219 (0.961)	0.049 (0.903)	0.268 ^{**} (0.293)	3.47	0.197 [*] (0.220)	2.55
WAIS-III Longest Digit Forward	-0.286 (0.956)	0.081 (0.894)	0.366 ^{***} (0.402)	5.05	0.373 ^{***} (0.422)	5.26
WAIS-III Longest Digit Backward	-0.352 (0.897)	0.090 (0.908)	0.441 ^{***} (0.487)	5.91	0.427 ^{***} (0.485)	5.87
Verbal Fluency	-0.349 (0.916)	0.121 (0.893)	0.470 ^{***} (0.522)	6.78	0.453 ^{***} (0.518)	6.69
Trails B	-0.257 (0.889)	0.094 (0.918)	0.351 ^{***} (0.386)	5.18	0.374 ^{***} (0.424)	5.65
Task efficiency						
Memory assessment measures						
Total recall	-0.587 (1.028)	0.109 (0.877)	0.696 ^{***} (0.777)	8.74	0.696 ^{***} (0.780)	8.87
Short-term memory	-0.447 (1.02)	0.062 (0.908)	0.509 ^{***} (0.563)	5.98	0.488 ^{***} (0.533)	5.80
Long-term memory	-0.455 (0.9863)	0.109 (0.895)	0.565 ^{***} (0.634)	7.57	0.558 ^{***} (0.618)	7.54
Attention & processing speed assessment measures						
Trails A	-0.486 (0.971)	0.090 (0.910)	0.575 ^{***} (0.628)	7.16	0.629 ^{***} (0.698)	7.93
CPT-II OMS	-0.375 (1.04)	0.046 (0.915)	0.420 ^{***} (0.458)	4.57	0.371 ^{***} (0.402)	4.06
CPT-II VAR	-0.304 (1.03)	0.064 (0.905)	0.368 ^{***} (0.402)	4.33	0.330 ^{***} (0.358)	4.16
WAIS-III Longest Digit Forward	-0.372 (0.961)	0.092 (0.916)	0.464 ^{***} (0.510)	6.27	0.460 ^{***} (0.504)	6.29
WAIS-III Longest Digit Backward	-0.448 (0.953)	0.101 (0.911)	0.548 ^{***} (0.595)	7.23	0.542 ^{***} (0.599)	7.24
Verbal Fluency	-0.466 (0.930)	0.148 (0.99)	0.614 ^{***} (0.676)	8.78	0.612 ^{***} (0.685)	8.87
Trails B	-0.340 (0.941)	0.126 (0.920)	0.465 ^{***} (0.502)	6.73	0.488 ^{***} (0.537)	7.16

		Impaired ^a	Not impaired ^a	Difference ^b (effect size) ^d	t	Difference ^c (effect size)	t
	Organization						
Memory assessment measures	Total recall	-0.189 (1.051)	0.043 (0.873)	0.232* (0.205)	2.59	0.230** (0.254)	2.88
	Short-term memory	-0.229 (1.050)	0.045 (0.877)	0.275*** (0.245)	2.93	0.260*** (0.287)	3.13
	Long-term memory	-0.135 (1.029)	0.039 (0.872)	0.173* (0.153)	2.15	0.164* (0.181)	2.20
Attention & processing speed assessment measures	Trails A	-0.135 (1.012)	0.026 (0.890)	0.160 (0.145)	1.86	0.190* (0.209)	2.38
	CPT-II OMS	-0.120 (1.054)	0.023 (0.881)	0.143 (0.128)	1.41	0.109 (0.121)	1.22
Executive functioning assessments	CPT-II VAR	-0.118 (1.014)	0.035 (0.876)	0.153 (0.138)	1.84	0.117 (0.129)	1.50
	WAIS-III Longest Digit Forward	-0.134 (0.991)	0.040 (0.887)	0.175 (0.192)	2.40	0.178* (0.196)	2.45
	WAIS-III Longest Digit Backward	-0.134 (0.973)	0.035 (0.896)	0.170 (0.186)	2.25	0.165* (0.181)	2.19
	Verbal Fluency	-0.141 (0.976)	0.050 (0.887)	0.192 (0.210)	2.73	0.182*** (0.200)	2.60
	Trails B	-0.021 (0.977)	-0.001 (0.887)	0.020 (0.022)	0.29	0.031 (0.034)	0.46
	Emotional regulation						
Memory assessment measures	Total recall	-0.362 (0.993)	0.068 (0.035)	0.430*** (0.462)	5.27	0.418*** (0.468)	5.32
	Short-term memory	-0.359 (0.961)	0.057 (0.927)	0.416*** (0.446)	4.86	0.374*** (0.417)	4.53
	Long-term memory	-0.310 (0.973)	0.078 (0.918)	0.388*** (0.417)	5.13	0.362*** (0.404)	4.93
Attention & processing speed assessment measures	Trails A	-0.215 (0.897)	0.038 (0.953)	0.254** (0.270)	3.09	0.311*** (0.344)	3.91
	CPT-II OMS	-0.363 (0.942)	0.056 (0.931)	0.419*** (0.449)	4.56	0.344*** (0.382)	3.85
	CPT-II VAR	-0.311 (0.974)	0.079 (0.918)	0.390*** (0.419)	4.96	0.331*** (0.368)	4.28
	WAIS-III Longest Digit Forward	-0.278 (0.951)	0.074 (0.927)	0.352*** (0.377)	4.73	0.341*** (0.380)	4.75
Executive functioning assessments	WAIS-III Longest Digit Backward	-0.279 (0.980)	0.064 (0.921)	0.342*** (0.366)	4.44	0.328*** (0.365)	4.42
	Verbal Fluency	-0.377 (0.953)	0.128 (0.904)	0.505*** (0.550)	7.15	0.504*** (0.573)	7.40
	Trails B	-0.179 (0.953)	0.065 (0.936)	0.243*** (0.258)	3.46	0.272*** (0.301)	4.02

Note:

^a Score of neurocognitive concerns estimated by IRT methods;

^b Unadjusted model;

^c Adjusted for current age, race, and gender;

^d Magnitude of effect size: <0.2 are classified as negligible, 0.2–0.49 as small, 0.5–0.79 as moderate, and >0.8 as large. Effect sizes 0.5 are considered as a clinically important difference;

* $p < 0.05$,

** $p < 0.01$,

*** $p < 0.001$.

Table 4
Correlation between revised latent CCSS-NCQ domain scores and continuous direct neurocognitive assessment scores

	Memory	Task efficiency	Organization	Emotional regulation
Memory assessment measures				
Total recall	0.34***	0.34***	0.13**	0.25***
Short-term memory	0.31***	0.28***	0.15**	0.23***
Long-term memory	0.30***	0.26***	0.13**	0.22***
Attention & processing speed assessment measures				
Trails A	0.16***	0.30***	0.09**	0.13***
CPT-II OMS ^a	-0.17***	-0.23***	-0.13**	-0.24***
CPT-II VAR ^a	-0.15***	-0.21***	-0.10**	-0.21***
Executive functioning assessments				
WAIS-III Longest Digit Forward	0.26***	0.28***	0.08**	0.24***
WAIS-III Longest Digit Backward	0.27***	0.33***	0.11**	0.21***
Verbal Fluency	0.27***	0.31***	0.09*	0.25***
Trails B	0.25***	0.34***	0.10**	0.21***

Note:

^a Higher scores indicate worse attention and processing speed;

* p<0.05,

** p<0.01,

*** p<0.001

Table 5
Multiple linear regression analysis of factors associated with neurocognitive concerns measured by the revised CCSS-NCQ

Variables	Memory		Task efficiency		Organization		Emotional regulation	
	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p
Age at diagnosis								
0-2	0.021	0.890	-0.039	0.803	-0.182	0.245	0.142	0.243
3-5	-0.082	0.532	0.042	0.752	-0.157	0.241	-0.003	0.980
6-10	-0.092	0.485	0.025	0.853	-0.170	0.206	0.045	0.672
11-15	-0.084	0.521	0.046	0.732	-0.194	0.148	0.021	0.838
16-20	Ref		Ref		Ref		Ref	
Cranial radiation								
Yes	-0.245	0.001	-0.333	<0.001	-0.033	0.655	-0.191	0.009
No	Ref		Ref		Ref		Ref	
Intravenous methotrexate								
Yes	-0.110	0.140	-0.056	0.467	-0.054	0.479	-0.078	0.296
No	Ref		Ref		Ref		Ref	
Intrathecal methotrexate								
Yes	0.032	0.731	-0.008	0.934	-0.049	0.599	-0.131	0.152
No	Ref		Ref		Ref		Ref	
High dose methotrexate								
Yes	0.102	0.213	0.160	0.058	0.102	0.221	0.184	0.028
No	Ref		Ref		Ref		Ref	
Relapse								
Yes	0.182	0.027	0.103	0.221	-0.050	0.553	0.198	0.017
No	Ref		Ref		Ref		Ref	
Current age								
17-24	0.178	0.141	-0.048	0.698	0.185	0.135	-0.003	0.983
25-29	0.095	0.392	-0.041	0.718	0.104	0.360	-0.065	0.619
30-34	0.107	0.314	0.089	0.419	0.117	0.283	-0.143	0.281
35-39	-0.057	0.575	0.038	0.717	0.030	0.774	-0.036	0.796
40+	Ref		Ref		Ref		Ref	

Variables	Memory		Task efficiency		Organization		Emotional regulation	
	Parameter estimate	<i>p</i>	Parameter estimate	<i>p</i>	Parameter estimate	<i>p</i>	Parameter estimate	<i>p</i>
Sex								
Female	-0.309	<0.001	-0.310	<0.001	-0.139	0.031	-0.467	<0.001
Male	Ref		Ref		Ref		Ref	
Race								
Other	0.103	0.247	0.094	0.305	-0.194	0.148	0.047	0.600
White	Ref		Ref		Ref		Ref	