

## RESEARCH ARTICLE

# Analyzing differences between parent- and self-report measures with a latent space approach

Dongyoung Go<sup>1,2</sup>, Minjeong Jeon<sup>3</sup>, Saebiyul Lee<sup>4,5</sup>, Ick Hoon Jin<sup>1,2\*</sup>, Hae-Jeong Park<sup>4,5,6\*</sup>

**1** Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea, **2** Department of Statistics and Data Science, Yonsei University, Seoul, Republic of Korea, **3** School of Education and Information Studies, University of California, California, Los Angeles, United States of America, **4** Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, Republic of Korea, **5** Department of Nuclear Medicine, Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea, **6** Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea

\* [parkhj@yonsei.ac.kr](mailto:parkhj@yonsei.ac.kr) (HJP); [ijin@yonsei.ac.kr](mailto:ijin@yonsei.ac.kr) (IHJ)



## OPEN ACCESS

**Citation:** Go D, Jeon M, Lee S, Jin IH, Park H-J (2022) Analyzing differences between parent- and self-report measures with a latent space approach. PLoS ONE 17(6): e0269376. <https://doi.org/10.1371/journal.pone.0269376>

**Editor:** James Mockridge, PLOS: Public Library of Science, UNITED KINGDOM

**Received:** June 30, 2021

**Accepted:** May 19, 2022

**Published:** June 29, 2022

**Copyright:** © 2022 Go et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data underlying the results presented in the study are available from [https://figshare.com/articles/dataset/lstrm\\_comparison\\_with\\_cbcl\\_ysr/19962830](https://figshare.com/articles/dataset/lstrm_comparison_with_cbcl_ysr/19962830).

**Funding:** This research was supported by Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C7A1030750), by the Yonsei University Research Fund of 2019-22-0210, and by Basic Science Research Program through the National Research Foundation of Korea (NRF 2020R1A2C1A01009881).

## Abstract

We explore potential cross-informant discrepancies between child- and parent-report measures with an example of the Child Behavior Checklist (CBCL) and the Youth Self Report (YSR), parent- and self-report measures on children's behavioral and emotional problems. We propose a new way of examining the parent- and child-report differences with an interaction map estimated using a Latent Space Item Response Model (LSIRM). The interaction map enables the investigation of the dependency between items, between respondents, and between items and respondents, which is not possible with the conventional approach. The LSIRM captures the differential positions of items and respondents in the latent spaces for CBCL and YSR and identifies the relationships between each respondent and item according to their dependent structures. The results suggest that the analysis of item response in the latent space using the LSIRM is beneficial in uncovering the differential structures embedded in the response data obtained from different perspectives in children and their parents. This study also argues that the differential hidden structures of children and parents' responses should be taken together to evaluate children's behavioral problems.

## Introduction

How children think of themselves may not be consistent with how their parents think about themselves. Reportedly, parents' evaluations of their children are often biased, sometimes superficial, and affected by their relationship with their children. Similarly, children often do not view themselves objectively [1–5].

Parent review and child review are frequently utilized in assessing children's behavior or psychology [6–9]. Among many scales, the Child Behavior Checklist (CBCL) [7] and the

**Competing interests:** The authors have declared that no competing interests exist.

Youth Self Report (YSR) are widely used scales to evaluate children's behavior problems, which are based on parent reports (CBCL) and self (children) reports (YSR) on the same questionnaires. To evaluate children, the CBCL and YSR comprise items of eight domains (often called syndromes): Aggressive Behavior (AB), Anxious/Depressed (AD), Attention Problems (AP), Rule-Breaking Behavior (RBB), Somatic Complaints (SC), Social Problems (SP), Thought Problems (TP), and Withdrawn/Depressed (WD). These domains are often categorized into scores of internalizing (a combination of AD, WD, and SC) and externalizing problems (a combination of RBB and AB).

The large discrepancies between the CBCL and YSR data are a commonly discussed issue in the literature [2, 6, 10–12]. To study the discrepancies between parent- and self-report measures from the CBCL and YSR, two types of approaches are conventionally taken: the direct comparison and the model-based approaches. For the direct comparison, researchers typically utilized the measures that estimate the dependency between items of two different scales, using Pearson correlation or Cohen's Kappa [2, 10, 12, 13]. However, these direct comparison methods do not take into account the potential dependency among items that may differently exist in the two sets of measures on children. Consequently, comparison between CBCL and YSR based on the responses without considering intrinsic item inter-dependency is questionable. Three types of inter-dependent hidden structures may exist in the group item response data; between items and items, between respondents and respondents, and between items and respondents. These inter-dependent structures may differ between CBCL and YSR.

To consider item-item inter-dependency, factor analysis [7, 14, 15] has been applied to evaluate the CBCL and YSR relationship by modeling item response data with linear combinations of latent factors. Not being based on the item-response theory, those latent variables do not directly reflect the innate item or respondent characteristics embedded in the choice behaviors.

Several studies exist that defined the characteristics of items based on the item-response theory model to compare the characteristics between the two groups of respondents in the latent space [2, 12, 16–19]. Although the item-response theory model includes latent variables for items and respondents to explain group response data (see Eq 2 in the Method section), the direct item-item, respondent-respondent, or item-respondent relationships are not easily discernible. Furthermore, the previous item-response theory model may not directly evaluate each individual's tendency toward items. Evaluating item response characteristics in each individual is critical in the practical application. Looking at individual-level discrepancies between a child-parent couple enables to generate personalized feedback on the child and parent. When exploring how a pair of child and parent views the child's behavior differently, it would be intuitive if the items and each respondent (child and parent) have their positions (thus can be visualized) in the same latent space and if the distance between each item and respondent is explicitly defined.

In this study, to explore the multi-dimensional inter-dependent structures between items, respondents and item-respondent pairs differently embedded in the CBCL and YSR responses, we approached the problem in the latent space with a newly developed Latent Space Item Response Model (LSIRM) [20].

Latent spaces, or in other words, interaction maps, are commonly utilized to represent the relationship or dependencies between actors in various literature, such as social network analysis [21, 22], recommendation [23, 24]. The LSIRM assumes the inter-dependency can be modeled with 'the distance' in latent space in which both respondents and items have their 'latent positions' where the short distances in the latent space implies the strong dependency. For example, two items closer to each other in a latent space have stronger dependency than two items far apart. Two items with strong dependency indicate that respondents tend to show a

similar response pattern to those items. Similarly, two respondents close to each other have strong dependencies, meaning that they tend to show a similar pattern of responding to test items (i.e., if one respondent responds positively to some items, the other respondent is likely to respond positively to those items). We will show that examining these dependency patterns in the two sets of measures can enlighten and shed new light on understanding the differences between CBCL and YSR. We will further present that the currently proposed method has an advantage in directly exploring respondent tendency toward items at the individual level, making it possible to examine different views in a child-parent pair on the child's behaviors.

The rest of the paper is organized as follows: We begin by describing the empirical data example that we will use in this study, followed by conventional analyses of the data. We then briefly describe the LSIRM and show how this model can capture the dependency in item response data. Next, we describe our strategy for comparing children and parent reports from the CBCL and YSR with LSIRM. In the Result and Discussion sections, we present and discuss our analysis results, compared with conventional analysis methods, and the differential views about the children's behaviors from the children's and their patients' perspectives.

## Materials and methods

### Empirical data example

**Data.** We used a dataset from the Children Mind Institute's healthy brain network MRI database (<https://childmind.org/data-sharing-initiatives/>) [25]. Among 1,479 children, we selected 662 children (male: 397, female: 265) who had both the CBCL and YSR scores. Their age ranged from 10–18 years (mean: 13.8 years, standard deviation: 1.97 years). The CBCL and YSR tests consist of 120 items stating identical children's behavioral problems in both tests. Among them, 118 items are multiple-choice questions such as “Acts too young for his/her age.” Respondents are asked to choose an option among the three response categories, “Not true,” “Somewhat or sometimes true,” “Very true or often true”, which are coded 0, 1, and 2, respectively. The other two free response items such as “Please write in any problems that were not listed above” is not considered here. The CBCL and YSR items are categorized into eight syndromes. We provided each item's syndrome membership in the [S1 Table](#). Several items do not belong to any particular syndromes. Those include Items 6, 7, 15, 24, 44, 49, 53, 55, 59, 60, 73, 74, 77, 80, 88, 92, 93, 98, 106, 107, 108, 109, and 110.

For data analysis, we dichotomized the original responses by combining the two positive categories of “Somewhat or sometimes true,” “Very true or often true,” and contrasting it with the only negative category of “Not true.” Such data dichotomization is not uncommon in the CBCL and YSR analysis, in part because of the frequently reported low reliability of the response categories [26–29]. After dichotomization, the mean proportion of positive responses was 0.43 with a standard deviation (SD) of 0.24 in the YSR, while in the CBCL, it was 0.27 with an SD of 0.17.

**Analysis of CBCL and YSR differences using conventional methods.** To compare with the current LSIRM analysis, we applied four analysis methods commonly used in the literature to compare the CBCL and YSR: three direct comparisons using pairwise measures, i.e., Pearson correlation, Kappa coefficients, and Jaccard similarity, and one model-based approach using item factor analysis. First, we evaluated the Pearson correlations between the CBCL and YSR at the syndrome level, using the syndrome-specific sum scores. The syndrome-level sum score is the sum of binary responses to the individual items, which can be treated as continuous data. Overall, the Pearson correlation was low, ranging from 0.23 (thought problem; TP) to 0.47 (rule-breaking problem; RBB). [Table 1](#) shows the CBCL-YSR correlations for all syndromes. The size of the correlations was similar to the reports in the literature (e.g., [13]).

**Table 1. Syndrome-level correlations and Kappa coefficients between CBCL and YSR.**

Syndrome	Correlation	Kappa
AP	0.30	0.20
RBB	0.47	0.28
AB	0.40	0.20
WD	0.38	0.22
TP	0.23	0.14
SP	0.27	0.18
AD	0.37	0.21
SC	0.31	0.18

<sup>a</sup>For the Kappa coefficients, we computed the syndrome mean of the item-level coefficients.

<https://doi.org/10.1371/journal.pone.0269376.t001>

We then computed the Kappa coefficients to evaluate the degree of agreement between CBCL and YSR at the item level, following the procedure used in [12]. The Kappa coefficients were also low, ranging from -0.02 to 0.49, with a mean of 0.18. Table 1 lists the mean Kappa coefficients for all syndromes. This result is also in line with the reports in the literature (e.g., [12]).

Next, we applied conventional dyadic similarity analysis at the item level by using the Jaccard similarity measure [30]. Jaccard similarity is a dyadic similarity measure that compares the positive response counts of item pairs. Specifically, Jaccard similarity for binary vectors  $A$  and  $B$  is defined as  $J(A, B) = |A \cap B| / |A \cup B|$ . For the CBCL and YSR data, we computed the Jaccard similarity between  $P$  item vectors in the CBCL and YSR, respectively, where  $P$  are the number of items; we then subtracted the CBCL Jaccard similarity from the YSR similarity matrix. Table 2 lists the item pairs with the top 12 most significant differences in Jaccard similarity. Interestingly, none of the identified items have syndrome membership.

Lastly, we applied exploratory item factor analysis to the CBCL and YSR data, using the R package ‘mirt’ [31]. In the item factor model, the probability of answering positively to item  $i$

**Table 2. Twelve item-item pairs are showing the most significant differences in Jaccard similarity between the CBCL and YSR.** [] indicates syndromes that the items belong to. [X] indicates that the item is not a member of any particular syndrome. For all pairs, the YSR showed high similarity (mean 0.963), while the CBCL showed low similarity (mean 0.014).

	Paired items
pair 1	107. Wets self during the day [X] ↔ 59. Plays with own sex parts in public [X]
pair 2	109. Whining [X] ↔ 59. Plays with own sex parts in public [X]
pair 3	59. Plays with own sex parts in public [X] ↔ 15. Cruel to animals [X]
pair 4	107. Wets self during the day [X] ↔ 92. Talks or walks in sleep [X]
pair 5	92. Talks or walks in sleep [X] ↔ 59. Plays with own sex parts in public [X]
pair 6	109. Whining [X] ↔ 107. Wets self during the day [X]
pair 7	59. Plays with own sex parts in public [X] ↔ 6. Bowel movements outside toilet [X]
pair 8	98. Thumb-sucking [X] ↔ 59. Plays with own sex parts in public [X]
pair 9	107. Wets self during the day [X] ↔ 6. Bowel movements outside toilet [X]
pair 10	107. Wets self during the day [X] ↔ 98. Thumb-sucking [X]
pair 11	109. Whining [X] ↔ 98. Thumb-sucking [X]
pair 12	88. Sulks a lot [X] ↔ 59. Plays with own sex parts in public [X]

<https://doi.org/10.1371/journal.pone.0269376.t002>

for respondent  $k$  is defined as follows:

$$\text{logit}(\mathbb{P}(y_{ki} = 1) | \boldsymbol{\alpha}_i, \boldsymbol{\theta}_k, d_i) = \boldsymbol{\alpha}_i^T \boldsymbol{\theta}_k + d_i \quad (1)$$

where  $d_i$  is item intercept,  $\alpha_i$  is item-specific latent factor, and  $\theta_k$  is respondent-specific latent factor [31].

In this experiment, we found the optimal number of factors was four, both for the CBCL and YSR data, based on G2 goodness-of-fit statistics and their p-values [32]. The factor structure comparison table and their fit statistics are given in the Supplementary S2 Table. The factor structure, however, was different between the two datasets. To summarize, for the first two factors, the CBCL and YSR show similar loading structures: the first significant factor covers most items of the AD syndrome (internalizing syndrome), and the second important factor covers most items of the RBB syndrome, an externalizing syndrome. However, the last two factor structures are quite different between the CBCL and YSR. The third factor is loaded on attention, social and thought problems (AP, SP, and TP) syndromes in the CBCL, while in the YSR, it is loaded on SC and WD syndromes. The fourth factor is loaded on the items with no syndrome membership in the YSR, while in the CBCL, this factor is loaded on most of the AB syndrome. These results are also consistent with the literature [6, 14].

All four analyses point to that there are significant discrepancies between the CBCL and YSR data. We will later compare and discuss the proposed latent space approach's results compared with these conventional methods. These conventional methods work as a reference for validating the proposed algorithm and providing a rationale for applying the current model-based approach.

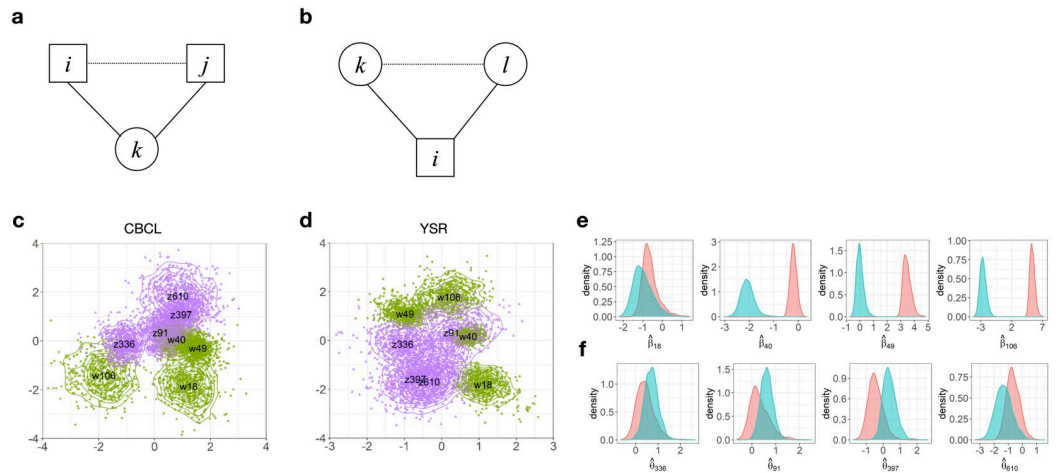
### Latent space item response model

The LSIRM [20] has been developed as an extension of the Rasch model to alleviate the conventional assumption of conditional independence (item responses are independent given a latent trait) and of homogeneity (respondents with the same trait level have equal success probability to an item). The LSIRM introduces latent positions,  $\mathbf{w}_i$  and  $\mathbf{z}_k$  for each item  $i$  and each respondent  $k$  in a  $d$ -dimensional latent space. The probability of positive response  $\mathbb{P}(y_{ki} = 1)$  to an item  $i$  from a respondent  $k$  is then formulated as follows:

$$\text{logit}(\mathbb{P}(y_{ki} = 1) | \beta_i, \theta_k, \mathbf{w}_i, \mathbf{z}_k) = \beta_i + \theta_k - \|\mathbf{w}_i - \mathbf{z}_k\|, \quad (2)$$

with item coefficients  $\beta_i$  (item easiness), respondent coefficients  $\theta_k$  (latent trait), and the latent positions of respondent  $\mathbf{z}_k$  and item  $\mathbf{w}_i$  ( $\mathbf{z}_k, \mathbf{w}_i \in \mathbb{R}^d$ ). The latent positions of respondent  $\mathbf{z}_k$  and item  $\mathbf{w}_i$  are determined by their distances, given the respondent and item coefficients. The resulting latent space approach provides an interaction map that represents the interactions of respondents and items, and helps derive insightful diagnostic information on items as well as respondents. We will use the words latent space and interaction map simultaneously. Though the dimension of the latent space or the distance measure can be arbitrarily chosen by researcher, we stick to  $\mathbb{R}^2$  and euclidean distance as Jeon et al. [20] remarked because of its strength in visualization. Eq 2 explains that when a respondent is further away from an item (i.e., larger distance and weaker dependency), the probability of giving a positive response to the item decreases. When a respondent becomes closer to an item (i.e., the shorter distance, the stronger dependency), the probability of giving a positive response increases. Note that Eq 2 is slightly different from Jeon et al. [20]'s formulation in that we fix the distance weight as one. This set-up enables us to match the two interaction maps from the CBCL and YSR data.

**Transitivity.** Note that although the model specifies distances between respondents and items in Eq 2, the model also captures item-item and respondent-respondent distances.



**Fig 1. The concept of the distance, position and the latent space (interaction map).** (a) and (b) visualize the triangle inequality of the distance. (a)  $k$  is a respondent, and  $i$  and  $j$  are two items. If the positions of two items  $i$  and  $j$  are close to respondent  $k$  in the interaction map, the two items are fairly close to each other, by the triangle inequality. (b)  $i$  is an item, and  $k$  and  $l$  are two respondents. If two respondents  $k$  and  $l$  are close to the item  $i$ , the two respondents  $k$  and  $l$  are also close. (c) is an example of the posterior distributions of latent positions in the CBCL interaction map, and (d) is the posterior distributions in the YSR interaction map. For visualization, only four items (w18, w40, w49, w106) and four respondents (z91, z336, z397, z610) are displayed. Green dots and lines around the item positions indicate the posterior distributions of the item positions. Purple dots and lines around the person positions indicate the posterior distributions of the person positions. Note that the interaction maps differ for parents (CBCL) and children (YSR). Note also that both item and person positions are located in the same map. The short distance between two latent positions of items (respondents) in the interaction map implies a large dependency between two items (respondents). (e) and (f) present posterior distributions of selected item coefficients (e) and respondent coefficients (f). Blue indicates the CBCL, and red indicates the YSR distributions for each coefficient. All Bayesian inferences are made through these posterior samples; thereby flexible inference such as variance or overlapped portion with other distribution can be made.

<https://doi.org/10.1371/journal.pone.0269376.g001>

Fig 1(a) and 1(b) explains how it is possible; (a) if a  $k$ -th respondent responds positively to an  $i$ -th item and at the same time responds positive to a  $j$ -th item, the latent positions of  $w_i$  and  $w_j$  are likely to be close to each other; (b) if an  $i$ -th item is positively answered by respondents  $k$  and  $l$ ,  $z_k$  and  $z_l$  are likely to be close to each other in interaction map. This is a property interaction map referred to as transitivity [33, 34].

**Dependency.** The distance between two latent positions in latent space is the measure of dependency. The short distance between  $i$ th item latent position ( $w_i$ ) and  $j$ th item latent position ( $w_j$ ) means high dependency between two items. The respondent who responds positively to item  $i$  is likely to respond positively to  $j$ th item, vice-versa. The LSIRM models the item-item, respondent-respondent, and item-respondent dependency by projecting respondents and items to the continuous latent space. Therefore, the model-based comparison of two scales from different informants using the LSIRM allows us to evaluate pairwise dependencies between items, between respondents, and between items and respondents. Additionally, comparisons in aspects of higher-order dependencies, such as clustering patterns and relationships between clusters, are available using the LSIRM.

Fig 1(c) and 1(d) illustrate the interaction maps of the CBCL and YSR for the selected respondents and items for specific example. The figures display the posterior samples of latent positions for the selected items and respondents. For the item-item comparison, Items 40 and 49 (w40, w49) are relatively close in both interaction maps. This implies that Items 40 and 49 have a high dependency in both YSR and CBCL responses. On the other hand, Items 49 and 106 (w49, w106) are apart more in the CBCL interaction map than in the YSR interaction



map, which implies a difference in dependency between the YSR and CBCL responses. Note that Item 40 states “Hears sounds or voices that aren’t there”, Item 49 states “Constipated, doesn’t move bowels”, and Item 106 is about “Vandalism”. Therefore, we can identify the differences of inherent dependency between pairs of items by comparing the interaction maps of the CBCL and YSR. For example, parents tended to make different reports on vandalism and constipation, while children responded similarly to those topics.

Regarding the respondent-respondent relationship, Fig 1(c) and 1(d) also show that some respondents are close to their locations, while others are further away. For example, Respondents 610 and 397 (z610, z397) are close to each other in both interaction maps, meaning that their response patterns were similar, given their latent trait levels and item difficulty. Similarly, for the respondent-item, we can also observe that Respondent 236 (z236) is closer to Item 106 (w106) in the CBCL space than in the YSR space. This indicates that for Respondent 236, the child’s probability of positively responding to “Vandalism” was lower than her parent’s likelihood of giving positive responses.

Of note, the current model assumes some hidden tendencies in the parents’ responses toward their children. Thus, other parent respondents’ responses are needed to decompose those hidden parent-to-children tendencies as a structure, even though the parent respondents are not directly related to the other children.

**Estimation.** To estimate the LSIRM parameters, we used a Bayesian approach, following [20]. The priors of  $\beta_p$ ,  $\theta_k$ ,  $\mathbf{z}_k$  and  $\mathbf{w}_i$  are set to an independent normal distribution with a mean of 0 and some variances. The hyper-prior for the variance parameter of  $\theta_k$  ( $k = 1, \dots, N$ ) is set to be a conjugate inverse Gamma distribution:

$$\begin{aligned}\beta_i | \tau_\beta^2 &\sim N(0, \tau_\beta^2), \\ \theta_k | \sigma^2 &\sim N(0, \sigma^2), \\ \sigma^2 &\sim \text{Inv} - \text{Gamma}(a_\sigma, b_\sigma), a_\sigma > 0, b_\sigma > 0, \\ \mathbf{w}_i &\sim \text{MVN}_d(\mathbf{0}, \mathbf{I}_d), \\ \mathbf{z}_k &\sim \text{MVN}_d(\mathbf{0}, \mathbf{I}_d).\end{aligned}$$

Metropolis-Hasting-within-Gibbs sampler [35] was used, which generates the posterior samples of  $\beta_p$ ,  $\theta_k$ ,  $\mathbf{z}_k$ , and  $\mathbf{w}_i$  for  $k = 1, \dots, N$  and  $i = 1, \dots, P$ . Additional details of the estimation procedure are provided in the S1 Appendix. We set the number of iterations to 30,000 and take every 5-th sample after the first 5,000 steps as a burn-in period. Convergence was satisfactory, and the posterior distributions of the example item and person parameters are presented in Fig 1(e) and 1(f).

Due to the invariance property of the distances (invariance to rotation, reflection, and translation), multiple solutions may be available for the latent positions that produce the same distance matrix. This issue is common for models that involve latent spaces [21]. To resolve such an identifiability issue, we applied Procrustes transformation [36] as a post-processing procedure, which is a standard method to resolve latent position identifiability in the literature [21, 22]. For  $[Z]$  the class of positions equivalent to  $Z$  under rotation, reflection, translation, the Procrustean transformation is  $Z^* = \text{argmin}_{TZ} \text{tr}(Z_0 - TZ)^T (Z_0 - TZ)$ , where  $Z_0$  is a fixed set of positions and  $T$  ranges over the set of rotations, reflections, and translations. It is known that  $Z^*$  is the closest element to  $Z_0$  in terms of the sum of squared and is unique if  $Z_0 Z^T$  is non-singular [37]. Here the target  $Z_0$  would be the positions draw that produces maximum a posteriori and the other posterior draw of latent positions are carried out Procrustes transformation to resolve the invariance to reflections, rotations and translations. All Bayesian inferences are made through the posterior samples obtained after the post-processing.

To deal with missingness in the datasets under investigation, we extended the estimation procedure of Jeon et al. [20] with Bayesian data augmentation [38]. Specifically, we imputed the missing data with the posterior samples, where an estimated item response for missing  $\hat{y}_{ik}$  was generated using the estimated parameters of the previous step of the Gibbs sampling. The imputed response values were updated in every Gibbs sampling step. After assuming missing at random (MAR), this Bayesian data augmentation produces valid imputation results [39]. Of all item and respondent pairs, the missing proportion was 0.192% in YSR and 0.259% in the CBCL. We made MAR assumption because currently there is no established method for testing the nature of missingness, and MAR is often assumed in the CBCL and YSR analysis. Further, addressing potential non-ignorable missingness is not the primary purpose of the current study. We will look into the sensitivity of this assumption in future research.

### Strategies for analyzing the CBCL and YSR differences with the LSIRM

We applied a series of further analyses for model-based examinations of the CBCL-YSR differences with the discussed latent space approach.

First, we fit the separate LSIRM to the CBCL and YSR datasets and estimated two sets of item coefficients  $\beta_i$ , respondent coefficients  $\theta_k$ , and their latent space positions ( $z_k$  and  $w_i$ ). We evaluated the goodness of fit of the model to each dataset.

Second, we examined the differences between the item coefficients  $\beta_i$  estimated from the CBCL and YSR analysis. We also compared item positions  $w_i$  in the CBCL and YSR interaction maps and checked whether the items with a large discrepancy between CBCL and YSR in item coefficients  $\beta_i$  show distinct interaction patterns between them.

Third, we compared the CBCL and YSR in terms of item-pair distance or item-item dependency over the interaction map. For this, we first evaluated the distance distribution of the two items  $i$  and  $j$  (i.e., the posterior distributions of the distance  $\|w_i - w_j\|$ ) in the CBCL denoted as  $D_{ij}^{CBCL}(x)$  and YSR denoted as  $D_{ij}^{YSR}(x)$ , respectively. This posterior distribution function  $D_{ij}^{CBCL}$  and  $D_{ij}^{YSR}$  were obtained by calculating the Euclidean distance of each posterior sample of  $w_i$  and  $w_j$  in CBCL and YSR, obtained by fitting the LSIRM. This distribution represents the estimated dependency between item  $i$  and item  $j$  in terms of the distribution. We then assessed the significance of the differences of each item-pair dependency by evaluating the overlapped portion of their distance distribution. The overlapped portion was defined as

$$R = \int_{-\infty}^{\infty} \min(D_{ij}^{CBCL}(x), D_{ij}^{YSR}(x)) dx,$$

where  $\min(D_{ij}^{CBCL}(x), D_{ij}^{YSR}(x))$  indicates the overlapping areas of the two distributions.  $R \leq .05$  indicates that the two distributions are fairly different in a statistical sense [40–42]. An alternative approach, the calculation of the Kullback-Leibler (KL) divergence between  $D_{ij}^{CBCL}(x)$  and  $D_{ij}^{YSR}(x)$ , was also applied to capture the difference of item-pair dependency and its results were reported in the Supplementary S2 Appendix.

Fourth, item syndrome-level dependency is evaluated, with the syndrome positions identified with the centroid of the syndrome-specific items. Since the item syndrome is named with its semantic properties, it makes the axis of the interaction map more interpretable and comparable with intuition. To highlight the meaning of this axis, we used cosine similarity to measure the similarity of item syndrome level. The cosine similarity is measured based on the coordinates of the syndrome latent positions in the CBCL and YSR separately. The cosine



similarity of two vectors  $\mathbf{a}$  and  $\mathbf{b}$  can be computed as

$$\cos(\phi) = \frac{\mathbf{a}^T \mathbf{b}}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2},$$

where  $\phi$  is the angle between two vectors  $\mathbf{a}$  and  $\mathbf{b}$ , and  $\|\mathbf{a}\|_2 > 0$  and  $\|\mathbf{b}\|_2 > 0$  are their lengths.  $\cos(\phi)$  ranges from -1 to 1, while -1 indicates that the two vectors point to the opposite directions, and 1 indicates the same direction. Cosine similarity of 0 means that the two vectors are orthogonal.

Fifth, we evaluated differences in the dependency between respondents with their syndrome levels. We presented that the distance between each respondent and syndrome in the interaction map could be used to find the vulnerable syndromes for each respondent. We then compared K-means clustering of all respondents based on the distance from all item syndromes and demonstrated how the interaction map approach could derive unique findings.

## Results

To evaluate the goodness of fit of the LSIRM to CBCL and YSR data, we assessed the proximity of the predicted item responses based on the estimated model to the original item responses, which is a commonly used strategy for evaluating the prediction accuracy of binary classification [43–45]. As evaluation criteria, we used sensitivity, specificity, and overall accuracy, which are reported as reliable measures [46, 47].

The three indices are defined as Specificity =  $\frac{TN}{TN+FP}$ , Sensitivity =  $\frac{TP}{TP+FN}$ , Overall accuracy =  $\frac{TP+TN}{TP+TN+FP+FN}$ , where TN is true negative, TP is true positive, FP is false positive, and FN is false negative. Table 3 shows the result. All values are higher than 0.70, except for the sensitivity for the CBCL. This suggests that overall, the LSIRM showed satisfactory fit to both the CBCL and YSR data in terms of prediction.

### Item coefficients and positions

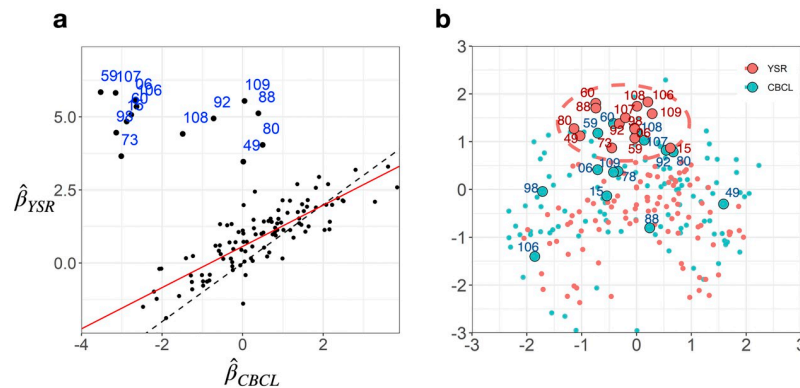
Fig 2(a) shows the relationship between the CBCL and YSR in the estimated item coefficients  $\hat{\beta}_i$ . For most test items, the two sets of  $\hat{\beta}_i$  are similar with the correlation of  $r = 0.83$ . However, a group of items does not follow this general pattern (marked in blue in the scatter plot). For those items,  $\hat{\beta}_{YSR}$  were higher than  $\hat{\beta}_{CBCL}$ , meaning that the children than their parents more easily endorsed them. The specific contents of those items were listed in Table 4. Those items appear to address behaviors related to sexual or physiological problems. That is, parents tended to believe that their children did not have sexual and physiological problems, even when the children themselves acknowledged such problems. Further analysis of these items with the interaction map can lead to more detailed interpretation.

Fig 2(b) shows interaction maps with item latent positions  $\mathbf{w}_i$  obtained from CBCL and YSR analysis with respective LSIRMs. For visual comparisons, the two estimated interaction maps were matched and integrated using Procrustes matching. Blue dots indicate the positions

**Table 3. Three goodness-of-fit measures of the LSIRM to the CBCL and YSR data.**

	CBCL	YSR
Specificity	0.927	0.846
Sensitivity	0.556	0.720
Overall accuracy	0.881	0.876

<https://doi.org/10.1371/journal.pone.0269376.t003>



**Fig 2. Item-wise comparison between the CBCL and YSR.** (a) Comparison between the CBCL and YSR item coefficients  $\hat{\beta}_i$  is displayed. The red line is the linear regression line between the CBCL and YSR estimates. The dotted line is  $y = x$ . Numbers indicate the outliers that deviate largely from the linear trend. For most test items, the two sets of  $\hat{\beta}_i$  are similar with the correlation of  $r = 0.83$ . However, the indicated items do not follow this general pattern. (b) An integrated interaction map for the CBCL and YSR item positions is presented. Blue and red dots represent the item positions identified in the CBCL and YSR data, respectively. Larger dots with number labels are the outlier items identified in (a). Note that the distance between dots indicates the degree of association of the two items in the latent space. For the close items, if respondents respond positively to an item, they are more likely to respond positively to the other item.

<https://doi.org/10.1371/journal.pone.0269376.g002>

of test items from the CBCL space, and red dots indicate the positions of test items in the YSR space. Respondents' positions are not presented in Fig 2(b) to focus on item position comparisons.

The distributions of item positions were generally similar between CBCL and YSR spaces, but there were some notable exceptions. The aforementioned items with dissimilar item coefficients (items in Fig 2(a) and Table 4) were placed in different regions of the CBCL and YSR interaction maps. See items marked with numbers in Fig 2(b). These items were closely located to each other in YSR (marked with a dotted red oval), which was not the case in the CBCL data.

### Differences in item-item dependency

We evaluated item-pair distance differences between the CBCL and YSR data based on  $D_{ij}^{CBCL}$  and  $D_{ij}^{YSR}$  values. For most item pairs, the difference was statistically significant ( $p < 0.01$ ) based on the Kolmogorov-Smirnov test. Table 5 lists the top 12 item pairs with the largest CBCL-YSR differences.

**Table 4. Items that differ between the CBCL and YSR in item-wise coefficients  $\hat{\beta}_i$  or item easiness (a tendency of the positive answer).** The item easinesses of the YSR for these items are greater than the CBCL ( $\hat{\beta}_{CBCL} < \hat{\beta}_{YSR}$ ).

Questions	
6. Bowel movements outside toilet	88. Sulks a lot
15. Cruel to animals	92. Talks or walks in sleep
49. Constipated, doesn't move bowels	98. Thumb-sucking
59. Plays with own sex parts in public	106. Vandalism
60. Plays with own sex parts too much	107. Wets self during the day
73. Sexual problems	108. Wets the bed
80. Stares blankly	109. Whining

<https://doi.org/10.1371/journal.pone.0269376.t004>

**Table 5. Top 12 item pairs that show the largest differences between the CBCL and YSR in terms of  $D_{ij}^{CBCL}$  and  $D_{ij}^{YSR}$ .** AP, SP and TP etc. within the brackets indicate the syndrome that each item belongs.

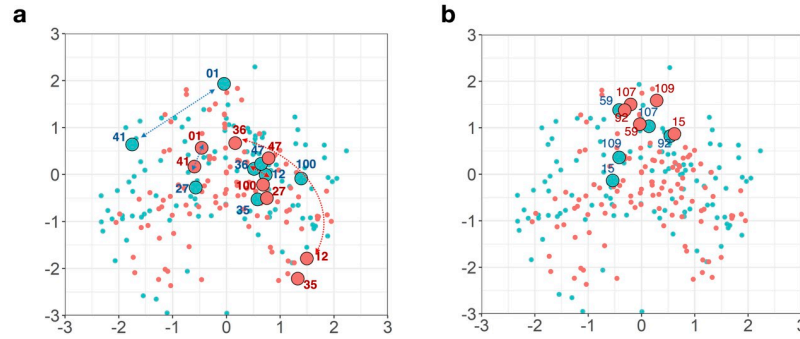
Pairs	Items	Distance (MAP)
Pair 1	1. Acts too young for his/her age [AP]	YSR (0.472) < CBCL (2.113)
	↔ 41. Impulsive or acts without thinking [AP]	
Pair 2	100. Trouble sleeping [TP]	YSR (0.342) < CBCL (1.938)
	↔ 27. Easily jealous [SP]	
Pair 3	100. Trouble sleeping [TP]	YSR (0.745) < CBCL (3.109)
	↔ 10. Can't sit still, restless or hyperactive [AP]	
Pair 4	100. Trouble sleeping [TP]	YSR (1.282) < CBCL (3.17)
	↔ 41. Impulsive or acts without thinking [AP]	
Pair 5	100. Trouble sleeping [TP]	YSR (0.752) < CBCL (2.964)
	↔ 78. Inattentive or easily distracted [AP]	
Pair 6	100. Trouble sleeping [TP]	YSR (0.627) < CBCL (2.535)
	↔ 79. Speech problem [SP]	
Pair 7	11. Clings to adults or too dependent [SP]	YSR (0.327) < CBCL (2.788)
	↔ 10. Can't sit still, restless or hyperactive [AP]	
Pair 8	11. Clings to adults or too dependent [SP]	YSR (0.841) < CBCL (3.009)
	↔ 41. Impulsive or acts without thinking [AP]	
Pair 9	11. Clings to adults or too dependent [SP]	YSR (0.338) < CBCL (2.462)
	↔ 78. Inattentive or easily distracted [AP]	
Pair 10	36. Gets hurt a lot, accident-prone [SP]	YSR (2.641) > CBCL (0.346)
	↔ 12. Complains of loneliness [SP]	
Pair 11	12. Complains of loneliness [SP]	YSR (2.451) > CBCL (0.534)
	↔ 83. Stores up too many things he/she doesn't need [TP]	
Pair 12	35. Feels worthless or inferior [AD]	YSR (2.588) > CBCL (0.809)
	↔ 47. Nightmares [SC]	

<https://doi.org/10.1371/journal.pone.0269376.t005>

Fig 3(a) illustrates how different the item-pair distances were between the CBCL and YSR data for the top four pairs. Interestingly, these items did not match the items identified from the item coefficient ( $\beta_i$ ) difference analysis (Table 4). In addition, they did not match the items identified with the Jaccard similarity analysis (Table 2 and Fig 3(b)). Fig 3(b) shows the Jaccard similarity between the CBCL and YSR data. The items identified in conventional direct comparison using Jaccard similarity were not coherently located with Fig 3(a) in the interaction map. This implies that our analysis of item-pair dependency does not offer the same kind of information as the analysis of the item coefficients and the conventional similarity analysis.

With item-pair differences, two groups of pairs can be distinguished in Table 5. In one group, those pairs were believed to have similar behaviors from the children's perspective, but not so much to the parents' view (Pairs 1–9). In the other group, the item pairs showed the opposite patterns: while they were perceived similar behaviors to the parents' eyes, they were believed to have different problems to the children (Pairs 10–12).

For example, to the children, "Act too young" and "Impulsive" (Pair 1) were similar behaviors, while it was not the case to the parents. The parents consider their children's behaviors of "Complains of loneliness" and "Stores up too many things he/she doesn't need" (Pair 11) are highly relevant, but children consider them irrelevant. Most of these items are the members of attention, social and thought problems (AP, SP, and TP) syndromes, indicating that those syndromes' behaviors were likely to be evaluated and reported differently by different informants, i.e., parents and children in the current context.

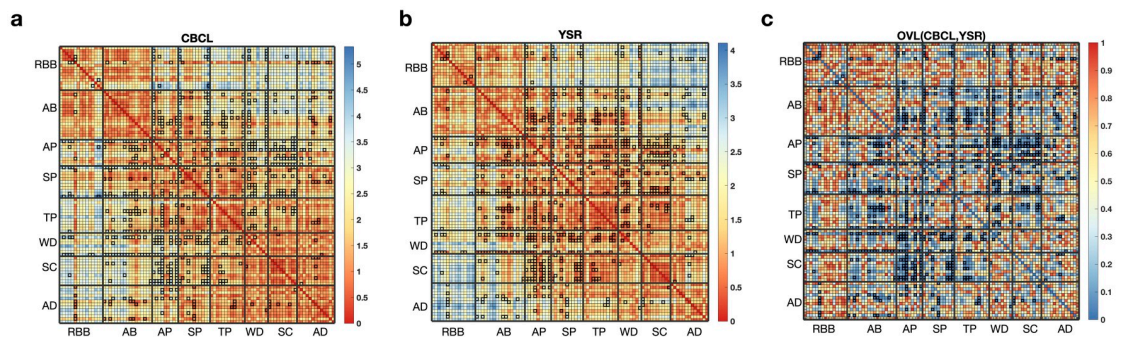


**Fig 3.** (a) Four item-item pairs with the largest differences in the distance (see Table 5). Double-side arrows are drawn to show the distances between two pairs (items 40 and 01; items 36 and 12) in the CBCL (blue) and YSR data (red). In addition to the position of each item, the distance between the two items' positions should be noted. (b) Four item-item pairs with the largest difference in the Jaccard similarity between the CBCL and YSR data. This implies that our analysis of item-pair dependency does not offer the same kind of information as the analysis of the item coefficient and the conventional similarity analysis.

<https://doi.org/10.1371/journal.pone.0269376.g003>

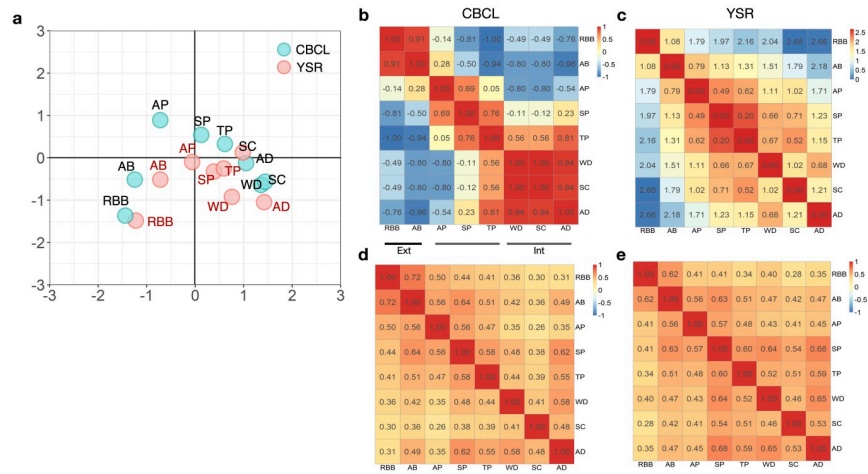
We assessed the overall differences in item-item dependency between the CBCL and YSR data by evaluating the overlap of the  $D_{ij}^{CBCL}$  and  $D_{ij}^{YSR}$  distributions. Fig 4 summarizes above Fig 3 with respect to item-item dependency categorized by syndromes. The distance of each item pair (i.e., expectation of the probability distribution  $D_{ij}$ ) is displayed in Fig 4(a) and 4(b) for the CBCL and YSR, respectively. The low value of distance means that their latent position was located closely and had a strong interaction. Fig 4(c) summarizes the overlapped portion for all pairs of items between CBCL and YSR. The item pairs with overlapped portion less than 0.05 are marked with bold edges. Of all the item pairs, about 9% of the pairs had overlapped portions of less than 0.05. The alternative approach using KL-divergence (S2 Appendix) suggests that pairs of items with a large discrepancy in KL-divergence between CBCL and YSR are positioned differently in their interaction maps.

**Syndrome-level distances.** Fig 5(a) shows the syndromes' latent positions (identified as the centroid of the syndrome item members) from the CBCL and YSR spaces. In both spaces,



**Fig 4.** Item-item distance heatmaps of the CBCL and YSR, and an overlapped portion heatmap between the CBCL and YSR. (a) displays the distance heatmap between pairs of item latent positions in the CBCL ( $D_{ij}^{CBCL}$ ) and (b) displays that in the YSR ( $D_{ij}^{YSR}$ ). (c) shows the heatmap of the overlapped portion between  $D_{ij}^{CBCL}$  and  $D_{ij}^{YSR}$  distributions for all item pairs. The item pair with a large overlapped portion means the perceived relation of that item pair is similar in both informants, while the item pair with a small overlapped portion means their relationship is perceived as different. The item pair with a large overlapped portion is colored with red, and the item pair with a small overlapped portion is colored with blue. The item pairs with overlapped portions less than 0.05 are marked with bold edges. About 9% of the pairs had overlapping portions of less than 0.05, and most of these pairs exist between different item syndromes (off-diagonal blocks), related to attention (AP), social (SP), and thought (TP) problems.

<https://doi.org/10.1371/journal.pone.0269376.g004>



**Fig 5. (a) Interaction map of the syndrome positions identified with CBCL (blue) and YSR (red).** Recall that syndromes are denoted with their abbreviation, Aggressive Behavior (AB), Anxious/Depressed (AD), Attention Problems (AP), Rule-Breaking Behavior (RBB), Somatic Complaints (SC), Social Problems (SP), Thought Problems (TP), and Withdrawn/Depressed (WD). These domains are often further categorized into internalizing (INT, a combination of AD, WD, and SC) and externalizing problems (EXT, a combination of RBB and AB). (b) and (c) are the heatmaps of the cosine similarity matrices of the item syndrome positions in the CBCL and YSR. (d) and (e) show the heatmaps of the correlation matrices of the raw syndrome scores in the CBCL and YSR. The identified dependency is roughly similar. However, the inter-correlations between externalizing syndromes (RBB, AB) and between internalizing syndromes (AD, WD, SC) were not outstanding in (d) and (e) compared to (b) and (c).

<https://doi.org/10.1371/journal.pone.0269376.g005>

externalizing syndromes (RBB, AB) are close to each other, and internalizing syndromes (AB, WD, and SC) are also close to each other, while externalizing and internalizing syndromes are far apart from each other. Fig 5(a) also shows that syndromes tend to be closer in their positions in the YSR space than in the CBCL space, meaning that the syndromes were seen more similar each other to the children’s perspective than to the parents’ view.

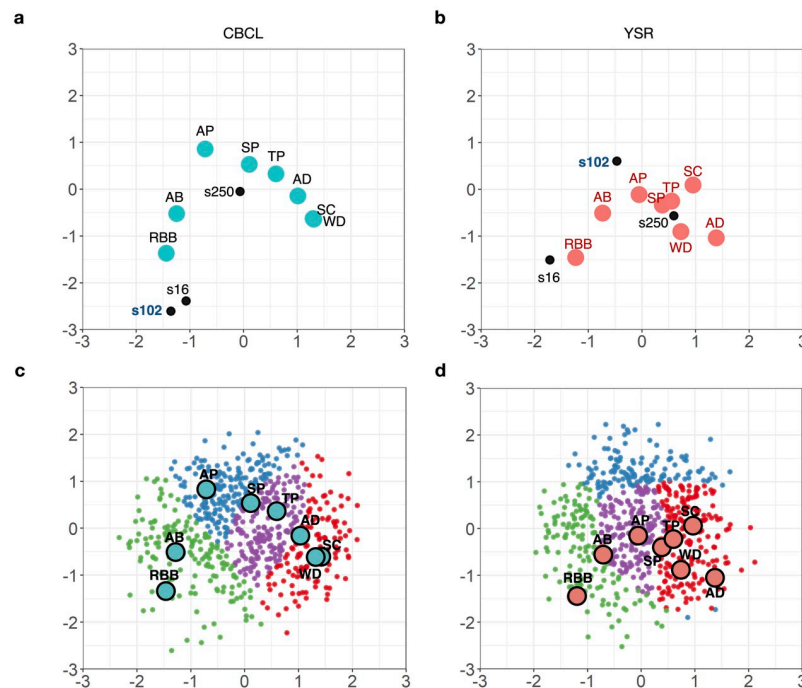
We measured cosine similarity between syndrome positions identified in the CBCL and YSR data. The cosine similarity is measured based on the coordinates of the syndrome latent positions in CBCL and YSR separately. Fig 5(b) and 5(c) display the heatmaps of cosine similarity measures for CBCL and YSR, and they suggest that the attention, social and thought problems (AP, SP, and TP) syndromes show different similarity patterns between CBCL and YSR. In particular, the characteristics of AP are distinct from other syndromes in the CBCL space. This result is consistent with the findings based on item-item dependency patterns.

In contrast, a simple inter-syndrome correlation analysis based on syndrome-level raw scores presented in Fig 5(d) and 5(e) did not identify inter- and intra-correlations between the syndromes. For example, the inter-correlations between externalizing syndromes (RBB, AB) and between internalizing syndromes (AD, WD, SC) were not outstanding.

### Respondent distances to syndromes

Individual respondents’ distances to the behaviors or the syndromes in the interaction map are useful as they are the basis to draw diagnostic conclusions about the respondents. To illustrate differences and similarities in the respondents’ distances to the syndromes, we selected three respondents as an example (s16, s102, and s250) in Fig 6(a) and 6(b). Respondent s16 was closest to the syndromes RBB and AB in both the CBCL and YSR spaces, indicating that she/he were more likely to show the behaviors of the RBB and AB syndrome than other





**Fig 6.** (a) and (b) display the syndrome positions and three respondents (s16, s102, s250) in the CBCL and YSR space, i.e., couples of parents (CBCL) and children (YSR), respectively. Note that both latent variables for items and latent variables for respondents are plotted in a latent coordinate. Latent positions for syndromes are presented as colored circles with solid black borders, and respondents' latent positions are displayed as dots. Three respondents are exemplary selected based on their latent positions in the CBCL and YSR: the two respondents located in the center (s250) and in the boundary (s16) show similarity between children and parents but s102 show difference in the latent positions between children and parents. (c) and (d) show the K-means clustering results of all respondents based on the distance from all item syndromes in the CBCL and YSR spaces.

<https://doi.org/10.1371/journal.pone.0269376.g006>

behaviors, and such a diagnosis of s16 would be similar based on the CBCL, and YSR reports. On the other hand, respondent s102 was located somewhat differently in the two spaces. This student was close to RBB and apart from AP in the CBCL space, but she/he was closer to AP than RBB in the YSR space.

Finding vulnerable syndromes using an interaction map is different from the conventional approach, such as simply comparing the positive response counts of syndromes for each respondent, because the latent position in the LSIRM is assigned by considering the dependency between all other objects. For example, one could draw the same conclusion with Fig 6(b) using simply comparing the positive response counts for s16, the respondent who answered about 2.5 times more positively to externalizing syndromes (23 positive responses) than internalizing syndromes (9 positive responses) in the YSR. On the other hand, s16 in the CBCL answered in externalizing and internalizing syndromes quite similarly (20 and 18 positive responses, respectively.) but is still located close to the RBB and AB, the externalizing syndromes in the interaction map, Fig 6(a). This is because the other respondents with similar response patterns to s16 in terms of each item, not discrete syndrome-level, responded more positively to externalizing syndromes. Therefore, although s16 responds similarly to both internalizing and externalizing syndromes in the CBCL, s16 might be more vulnerable to externalizing syndromes when considering other respondents with similar patterns. This means that one could end up with different conclusions about their most likely behaviors and



the existence of cross-informants discrepancy depending on whether the dependency was considered or not.

We applied K-means clustering to the respondent positions to see the differences in the respondent-syndrome distances between the CBCL and YSR spaces. The results are shown in Fig 6(c) and 6(d). Four clusters were identified, each with the CBCL and YSR respondent positions (distinguished with four colors). However, the four clusters identified from the two spaces do not necessarily match in terms of their most likely syndromes, given clusters' different distances to the syndromes. For example, the cluster located in the upper side of the interaction map (blue) did not appear close to any particular syndrome in the YSR space (Fig 6(d)), while such cluster did not exist in the CBCL space (Fig 6(c)). This re-confirms that using either parent's or children's reports can lead us to a different conclusion about the children's most likely syndromes and behavioral problems.

## Discussion

This paper discusses a new way of exploring the hidden structure of item responses from different informants with a latent space modeling approach, referred to as the LSIRM method. This new methodology was applied to the comparative analysis of the CBCL and YSR, which are widely-used multi-item scales on children's behavioral and emotional problems, based on parent- and self-reports, respectively. The LSIRM analysis spotted substantial differences between the CBCL and YSR measures which were not examined with existing methods, either direct comparison approaches or model-based item factor analysis.

Direct comparison analysis using Pearson correlation, Cohen's Kappa, or Jaccard similarity, computes the similarity between each item (or between each syndrome, with the sum of items belonging to it) pair-wisely and thereby treat each subject independently. This type of direct item-wise approach does not consider the potential dependency that may exist in other items or other respondents and cannot be used to compare the dependency between items and respondents at the same time. To figure out the item-respondent relationship, additional analysis such as linear regression or subject grouping is needed [2, 12, 18, 19]. We do not disregard the possibility that some of those dependencies can be explored by extending the conventional methods. For example, the residual covariance matrix may be used to find item-item dependencies. We have found that conventionally used direct comparison methods partly capture the relationship inherent in data more obtusely but sometimes lead to different conclusions (e.g., Fig 3). The consistency between the results of the LSIRM and direct comparison methods reported in the previous literature and this study supports the validity of the proposed algorithm. This validity is critical for the LSIRM, which has inherent overfitting problems as a complex model. It also lays the basis for the new findings' reliability derived from the further analysis unique to the LSIRM.

Item factor analysis [48, 49], shown in Eq 1, is designed to identify the latent factor structure underlying the item response data by assuming latent factors to be linear combinations of observed item responses. The item clusters identified with item factor analysis (S2 Table) are roughly similar to the LSIRM, showing high consistency between the CBCL and YSR in the internalizing and externalizing syndromes. That is, item factor analysis can identify clustering of items similar to the LSIRM. More specifically, the item factor analysis can capture item-by-respondent interactions to some degree in the sense that factor loadings (specific to items) are multiplied by factors (specific to respondents) (See Eq 1). Nevertheless, not all item-by-respondent interactions can be quantified with the item factor analysis because factor loadings are item-specific, not item-by-respondent specific. No parameters represent interactions between items and respondents in the item factor analysis model (Eq 1). Furthermore, the item factor

analysis cannot capture the similarity of respondents in terms of response patterns. In contrast, the LSIRM explicitly models item-by-respondent interactions in the form of distances in a metric space, which makes it possible to examine similarities among items and similarities among respondents (Eq 2). As exemplified in Fig 6, the characteristics of each individual can intuitively be represented in terms of the respondent's position and distances from items in the LSIRM. This model enables the investigation of each child and parent pair precisely. Of note, in terms of model parameter estimation, we used a Bayesian estimation with uninformative priors; thus, it is less likely that the model inversion scheme based on Bayesian or frequentist approaches may not be a significant factor in deriving the results.

The LSIRM differentiates between each item's coefficient and item position in the latent space. The item coefficient is estimated independent of other items and individual characteristics. In contrast, the item position is estimated by considering structured distances from respondents (Eq 2). The highly dependent structure of item response data differently embedded in the CBCL and YSR is first reflected in the item coefficient analysis. The high accordance exists in most item-wise coefficients between the CBCL and YSR, except for some outliers (Fig 2(a)). The items with a large discrepancy appear primarily related to sexual or physiological problems (Table 4). Parents do not respond positively to sexual and physiological issues of their children, even though their children are aware of these problems. This is predictable and is consistent with the findings in previous studies [50], which were revealed by using simple correlation analysis of the responses. In this respect, the current approach for analyzing item-wise coefficients in the latent space does not significantly differ from the direct comparison methods such as the Jaccard similarity measure (to compare the positive response count) in identifying discrepant items between the CBCL and YSR. However, the current study further characterizes the structured relationships between those items by presenting their positions in the latent space of the CBCL and YSR. As shown in Fig 2(b), those items are densely located on the upper boundary of the YSR, forming a clear cluster, but not in CBCL.

For items related to sexual or physiological problems with a large difference in item coefficients,  $\beta_i$ , the positive response count is more endorsed in children than parents. Thus, those items have a more clustering tendency in the YSR. That means the respondent who positively responded to one of these outlier items also showed similar response patterns to other outlier items in the YSR data. In contrast, those patterns were not shown in the CBCL data. In other words, children think those items are highly related to each other, which is not considered the same for parents. This item dependency pattern was more evidenced in the item-item distance analysis.

In addition to providing a method to compare two scales using each item's position, the advantage of the LSIRM-based approach is its ability to identify item-item distances in a structured way in the latent space. The item-item pairs that show differences between children and parents can be divided into two groups. One is the item pairs that children perceived as the underlying relationship while parents did not find them related to their children (Pairs 1–9 in Table 5). The other is item pairs that children did not find a relationship between domains in their behaviors, while parents observed a strong relationship between the two items in their children. For example, parents implicitly perceived the items in pair 1, i.e., “Act too young” and “Impulsive”, as not much relevance to children, while their children regarded the two as highly relevant. Furthermore, pairs 2–6 were related to “Trouble sleeping”. Trouble in sleeping may manifest in various syndromes, particularly in attention, such as inattentive or impulsive. However, it may be difficult for parents (as observers) to perceive the underlying connection. Meanwhile, parents consider that “Complains of loneliness” and “Stores up too many things he/she doesn't need” are highly relevant items regarding their children, but children consider those items are irrelevant from their perspectives on themselves.

Note that the top item pairs showing the largest difference between parents and children in terms of item-item distance (Table 5) belong mostly to pairs between non-extreme syndromes (attention, social and thought problems (AP, SP, and TP)). Only a few item pairs from the same syndrome showed differences between children and parents, indicating that the item-item similarity within the same syndromes was relatively consistent across children and parents. This is conspicuous for items in syndromes with extreme properties (i.e., externalizing scores, internalizing scores), as was found in previous studies that report high consistency in extreme problem scores [7, 10, 13, 51].

Of note, the item-item pairs that showed distance differences between the children's and parents' views on the children (Fig 3(a)) are composed of items that do not belong to the list of items showing item-wise parameter differences between the CBCL and YSR (Fig 2). Meanwhile, the conventional response-based approach using the Jaccard similarity did not provide additional information compared to item-wise analysis of the difference between the CBCL and YSR. The item pairs with the most considerable difference of the Jaccard similarity (Fig 3(b)) are almost combinations of the outlier items in Fig 2(a). Therefore, the conventional dyadic comparison did not separate item-specific effects and the interactions within the CBCL or YSR.

Overlapped portion analysis of items' latent positions between the CBCL and YSR also revealed similar results to the item-item distance analysis. Several item pairs in attention, social, and thought problems (AP, SP, and TP) have a significantly different relationship in the CBCL and YSR, especially for item pairs with one belonging to one of AP, SP, or TP and the other belonging to externalizing or internalizing scores Fig 4(c). In addition, most of the item pairs that showed high differences between children and parents were closer in the children compared to the parents (Fig 4).

Taking all results together, the current item-item dependency analysis using the LSIRM successfully specified the structural differences of children's self-views and parents' views on the children. Our latent space approach can reveal non-ignorable item-item dependence and identify that item-item dependence patterns and magnitude are fairly different between the CBCL and YSR data. In the CBCL-YSR context, where the test items indicate behavioral symptoms for certain syndromes, the identified difference in item dependency suggests that one could draw different conclusions, prevalent problems (or symptoms) and their relationships when data analysis was based on a single source, parent- or child- report only.

When we analyzed the latent position of the item syndrome, it shows directional order along the x-axis, i.e., the direction of 'external' to 'internal' axis (Fig 5(a)). This polarized property of externalizing and internalizing scores is consistent with the results of the item factor analysis [7]. Indeed, the axis of the fitted latent space does not have an interpretative meaning. However, by displaying the positions of item syndromes, one can make inferences about the space with reference to the position of each item syndrome. We found that the positions of syndromes in the YSR were generally closer than those in the CBCL. The congregation of the YSR syndromes reflects stronger interaction and a closer distance between syndromes in the YSR compared to the CBCL. This may be attributable to a tendency of positive answers in the YSR overall [51, 52]. The calculated cosine similarity (in Fig 5(b) and 5(c)) shows more apparent separation for items in externalizing and internalizing scores. Item syndromes belonging to externalizing or internalizing scores are generally perceived to be strongly related within the same group in both the CBCL and YSR, except for reduced association with SC (WD and SC) in the CBCL. However, for the attention, social, and thought problems (AP, SP, and TP), the perceived item relationship differs between the CBCL and YSR. This result is consistent with the findings based on the item-item distance pattern and their overlapped portion heatmap (Fig 4). Note that the distance among nonextreme syndromes (the attention, social, and

thought problems) differ obviously between children and parents. Thus categorization of items done in the parent's perspective may not be sufficient to characterize children's diverse problems and demands a new attempt to make a more specific division of questionnaires.

Because the latent position of each respondent is mapped with the items' latent position, clustering of respondents can identify not only a group of respondents who responded positively to the same item but also a group of respondents who responded positively to items with similar characteristics. This type of clustering index can be used to determine the relationship between each respondent or the discordance between parents and their children. For example, the clustering of respondents based on the distance between the item syndrome and each respondent (Fig 6(c) and 6(d)) shows that a cluster of the YSR located on the upper side of the latent space did not overlap with any specific symptoms and can be characterized by positive responses exclusively to items in Fig 2(b). This analysis revealed that the respondent-syndrome distances were not identical to some respondents in the CBCL and YSR spaces. The syndromes close to the respondent are the syndromes that the respondents (children in our context) are likely to have. Thus, this finding implies that one could draw different diagnostic conclusions about some children depending on whether children's or parents' reports were chosen for data analysis. These results also suggest that the conventional division of symptoms and clustering of individuals according to the dominant symptomatic problem may not necessarily apply to the division of the YSR.

Discrepancies between parent- and self-report measures have been commonly discussed, although examining and evaluating such discrepancies is an open area of research. The current study introduced a new and unique way of assessing parent- and self-report differences with a latent space approach using the LSIRM. Our empirical data analysis demonstrated that the two sets of measures from the two different informants did show much difference from the perspectives that have not been investigated with standard methods. Some of the differences that we identified have not been seen before and hold important practical implications. For example, behaviors and syndromes might show different structural relationships based on parent- or self-report measures. In addition, children might end up being diagnosed with a different set of behaviors and syndromes depending on whether parent- or self-report measure was used. Using the LSIRM, we mainly illustrated the potential discrepancies between parent- and child-report measures in terms of dependency. The identified CBCL-YSR differences from our proposed method illuminate and invite applied researchers to investigate and take a closer look at those discrepancies in other CBC-YSR datasets or other types of cross-informant measures.

In the current study, we did not examine how to resolve the identified differences, which is undoubtedly an important question, and we reserve it for future research. We primarily focused on the differences between informants in the latent space. Further analyses identifying the similarly responded items between informants, particularly items beyond internalizing and externalizing syndromes, remain as future research. Furthermore, the proposed model-based approach can be extended to a wide range of future work. First, our approach can be extended to non-binary item response data by choosing a suitable link function for non-binary item response data like in generalized linear models [20, 53]. Second, we can incorporate explanatory variables in our model to evaluate the discrepancy between the YSR and CBCL caused by the covariate information of the children, such as age [54], intelligence score, sex [1], culture, and ethnicity [11].

## Supporting information

**S1 Appendix. Metropolis-Hastings within Gibbs sampling for LSIRM.**  
(PDF)

**S2 Appendix. Item divergence using KL-divergence.**

(PDF)

**S1 Table. Items' syndrome membership.**

(PDF)

**S2 Table. Factor structures for CBCL and YSR.**

(PDF)

**Author Contributions****Conceptualization:** Ick Hoon Jin, Hae-Jeong Park.**Data curation:** Saebyul Lee, Hae-Jeong Park.**Formal analysis:** Ick Hoon Jin, Hae-Jeong Park.**Funding acquisition:** Ick Hoon Jin, Hae-Jeong Park.**Investigation:** Minjeong Jeon, Hae-Jeong Park.**Methodology:** Dongyoung Go, Ick Hoon Jin, Hae-Jeong Park.**Software:** Dongyoung Go, Ick Hoon Jin.**Supervision:** Saebyul Lee, Hae-Jeong Park.**Visualization:** Hae-Jeong Park.**Writing – original draft:** Dongyoung Go, Minjeong Jeon, Ick Hoon Jin, Hae-Jeong Park.**Writing – review & editing:** Hae-Jeong Park.**References**

1. Barrett M, Berney T, Bhate S, Famuyiwa O, Fundudis T, Kolvin I, et al. Diagnosing Childhood Depression. Who Should be Interviewed-Parent or Child? *The British Journal of Psychiatry*. 1991; p. 22–27. PMID: [1840754](https://pubmed.ncbi.nlm.nih.gov/1840754/)
2. Berg-Nielsen TS, Vika A, Dahl AA. When adolescents disagree with their mothers: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. *Child: Care, Health, and Development*. 2003; 29:207–213. <https://doi.org/10.1046/j.1365-2214.2003.00332.x> PMID: [12752611](https://pubmed.ncbi.nlm.nih.gov/12752611/)
3. Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*. 1991; 19:75–95. <https://doi.org/10.1007/BF00910566> PMID: [2030249](https://pubmed.ncbi.nlm.nih.gov/2030249/)
4. Martin JL, Ford CB, Dyer-Friedman J, Tang J, Huffman LC. Patterns of agreement between parent and child ratings of emotional and behavioral problems in an outpatient clinical setting: When children endorse more problems. *Journal of Developmental & Behavioral Pediatrics*. 2004; 25:150–155. <https://doi.org/10.1097/00004703-200406000-00002>
5. Youngstrom E, Loeber R, Stouthamer-Loeber M. Patterns and correlates of agreement between parent, teacher, and self-reported and parent-reported emotional and behavioral problems, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*. 2000; 68:1038. <https://doi.org/10.1037/0022-006X.68.6.1038> PMID: [11142538](https://pubmed.ncbi.nlm.nih.gov/11142538/)
6. Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Department of Psychiatry, University of Vermont; 1991. Available from: <https://books.google.co.kr/books?id=I5btOwAACAAJ>.
7. Achenbach Thomas M, Rescorla L. *Manual for the ASEBA Preschool forms and Profiles*. L.A.: Burlington, Vt.: ASEBA; 2002.
8. Sisteré ML, Massons JMD, Pérez RG, Ascaso LE. Validity of the DSM-oriented scales of the child behavior checklist and youth self-report. *Psicothema*. 2014; p. 364–371.
9. Bordin IA, Rocha MM, Paula CS, Teixeira MCT, Achenbach TM, Rescorla LA, et al. Child Behavior Checklist (CBCL), Youth Self-Report (YSR) and Teacher's Report Form (TRF): an overview of the

- development of the original and Brazilian versions. *Cadernos de Saúde Pública*. 2013; 29:13–28. PMID: [23370021](#)
10. Lacalle M, Ezpeleta L, Doménech JM. DSM-Oriented Scales of the Child Behavior Checklist and Youth Self-Report in Clinically Referred Spanish Children. *The Spanish Journal of Psychology*. 2012; 15:377–387. [https://doi.org/10.5209/rev\\_SJOP.2012.v15.n1.37344](https://doi.org/10.5209/rev_SJOP.2012.v15.n1.37344) PMID: [22379727](#)
  11. Rescorla LA, Ginzburg S, Achenbach TM, Ivanova MY, Almqvist F, Begovac I, et al. Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology*. 2013; 42:262–273. <https://doi.org/10.1080/15374416.2012.717870> PMID: [23009025](#)
  12. Gearing RE, Schwalbe CS, MacKenzie MJ, Brewer KB, Ibrahim RW. Assessment of adolescent mental health and behavioral problems in institutional care: discrepancies between staff-reported CBCL scores and adolescent-reported YSR scores. *Adm Policy Ment Health*. 2015; 42(3):279–287. <https://doi.org/10.1007/s10488-014-0568-y> PMID: [24938476](#)
  13. Sourander A, Helstelä L, Helenius H. Parent-adolescent agreement on emotional and behavioral problems. *Social psychiatry and psychiatric epidemiology*. 1999; 34:657–663. <https://doi.org/10.1007/s001270050189> PMID: [10703276](#)
  14. Hartman CA, Hox J, Auerbach J, Erol N, Fonseca AC, Mellenbergh GJ, et al. Syndrome dimensions of the Child Behavior Checklist and the Teacher Report Form: A critical empirical evaluation. *Journal of Child Psychology and Psychiatry*. 1999; 40:1095–1116. <https://doi.org/10.1111/1469-7610.00527> PMID: [10576539](#)
  15. de Groot A, Koot HM, Verhulst FC. Cross-cultural generalizability of the Youth Self-Report and Teacher's Report Form cross-informant syndromes. *Journal of Abnormal Child Psychology*. 1996; 24:651–664. <https://doi.org/10.1007/BF01670105> PMID: [8956089](#)
  16. Lambert MC, Essau CA, Schmitt N, Samms-Vaughan ME. Dimensionality and psychometric invariance of the Youth Self-Report Form of the Child Behavior Checklist in cross-national settings. *Assessment*. 2007; 14,3:231–45. <https://doi.org/10.1177/1073191107302036> PMID: [17690380](#)
  17. Panter A, Reeve BB. Assessing tobacco beliefs among youth using item response theory models. *Drug and Alcohol Dependence*. 2002; 68:21–39. [https://doi.org/10.1016/S0376-8716\(02\)00213-2](https://doi.org/10.1016/S0376-8716(02)00213-2) PMID: [12324173](#)
  18. Ferro MA, Boyle MH, Scott JG, Dingle K. The child behavior checklist and youth self-report in adolescents with epilepsy: testing measurement invariance of the attention and thought problems subscales. *Epilepsy & Behavior*. 2014; 31:34–42. <https://doi.org/10.1016/j.yebeh.2013.11.009> PMID: [24333500](#)
  19. Van Lieshout RJ, Boyle MH, Schmidt L, Saigal S, Ferro MA. Measurement Invariance Across Parent and Self-Ratings of Extremely Low Birth Weight Survivors and Normal Birth Weight Controls in Childhood and Adolescence on the Child Behavior Checklist and Youth Self-Report. *J Pediatr Psychol*. 2015; 40(8):825–833. <https://doi.org/10.1093/jpepsy/jsv043> PMID: [25981895](#)
  20. Jeon M, Jin IH, Schweinberger M, Baugh S. Mapping unobserved item-response interactions: A latent space item response model with interaction maps. *Psychometrika*. 2021;. <https://doi.org/10.1007/s11336-021-09762-5>
  21. Hoff P, Raftery A, Handcock MS. Latent space approaches to social network analysis. *Journal of the American Statistical Association*. 2002; 97:1090–1098. <https://doi.org/10.1198/016214502388618906>
  22. Sewell DK, Chen Y. Latent space models for dynamic networks. *Journal of the American Statistical Association*. 2015; 110:1646–1657. <https://doi.org/10.1080/01621459.2014.988214>
  23. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer*. 2009; 42(8):30–37 <https://doi.org/10.1109/MC.2009.263>
  24. Rendle S. Factorization machines. In: 2010 IEEE International conference on data mining. IEEE; 2010. p. 995–1000.
  25. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*. 2017; 4:1–26. <https://doi.org/10.1038/sdata.2017.181> PMID: [29257126](#)
  26. Pandolfi V, Magyar CI, Dill CA. An initial psychometric evaluation of the CBCL 6–18 in a sample of youth with autism spectrum disorders. *Research in Autism Spectrum Disorders*. 2012; 6:96–108. <https://doi.org/10.1016/j.rasd.2011.03.009> PMID: [22059091](#)
  27. Read KL, Settapani CA, Peterman J, Kendall PC, Compton S, Piacentini J, et al. Predicting anxiety diagnoses and severity with the CBCL-A: improvement relative to other CBCL scales? *Journal of Psychopathology and Behavioral Assessment*. 2015; 37:100–111. <https://doi.org/10.1007/s10862-014-9439-9> PMID: [26257470](#)



28. Sood BG, Bailey BN, Covington C, Sokol RJ, Ager J, Janisse J, et al. Gender and alcohol moderate caregiver reported child behavior after prenatal cocaine. *Neurotoxicology and Teratology*. 2005; 27:191–201. <https://doi.org/10.1016/j.ntt.2004.10.005> PMID: 15734270
29. Tan TX, Dedrick RF, Marfo K. Factor structure and clinical implications of child behavior checklist/1.5–5 ratings in a sample of girls adopted from China. *Journal of Pediatric Psychology*. 2007; 32:807–818. <https://doi.org/10.1093/jpepsy/jsm025> PMID: 17449465
30. Choi S, Cha SH, Tappert C. A Survey of Binary Similarity and Distance Measures. *J Syst Cybern Inf*. 2009; 8.
31. Chalmers RP. mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*. 2012; 48:1–29. <https://doi.org/10.18637/jss.v048.i06>
32. McKinley RL, Mills CN. A comparison of several goodness-of-fit statistics. *Applied Psychological Measurement*. 1985; 9(1):49–57. <https://doi.org/10.1177/014662168500900105>
33. Holland PW, Leinhardt S. Transitivity in structural models of small groups. *Comparative Group Studies*. 1971; 2:107–124. <https://doi.org/10.1177/104649647100200201>
34. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *nature*. 1998; 393:440–442. <https://doi.org/10.1038/30918> PMID: 9623998
35. Chib S, Greenberg E. Understanding the metropolis-hastings algorithm. *The American Statistician*. 1995; 49:327–335. <https://doi.org/10.1080/00031305.1995.10476177>
36. Gower JC. Generalized procrustes analysis. *Psychometrika*. 1975; 40:33–51. <https://doi.org/10.1007/BF02291478>
37. Sibson R. Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1979; 41(2):217–229.
38. Tanner MA, Wong WH. The Calculation of Posterior Distributions by Data Augmentation (with discussion). *Journal of the American Statistical Association*. 1987; 82:528–550. <https://doi.org/10.2307/2289463>
39. Little RJ, Rubin DB. *Statistical analysis with missing data*. vol. 793. John Wiley & Sons; 2019.
40. Pastore M, Calcagni A. Measuring distribution similarities between samples: a distribution-free overlapping index. *Frontiers in Psychology*. 2019; 10:1089. <https://doi.org/10.3389/fpsyg.2019.01089> PMID: 31231264
41. Queen JP, Quinn GP, Keough MJ. *Experimental design and data analysis for biologists*. Cambridge university press; 2002.
42. Neyman J, Pearson ES. The testing of statistical hypotheses in relation to probabilities a priori. In: *Mathematical proceedings of the Cambridge philosophical society*. vol. 29. Cambridge University Press; 1933. p. 492–510.
43. Yerushalmy J. Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports (1896-1970)*. 1947; p. 1432–1449. <https://doi.org/10.2307/4586294> PMID: 20340527
44. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006; 27:861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
45. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:201016061*. 2020.
46. Warnick EM, Bracken MB, Kasl S. Screening efficiency of the Child Behavior Checklist and Strengths and Difficulties Questionnaire: A systematic review. *Child and Adolescent Mental Health*. 2008; 13:140–147. <https://doi.org/10.1111/j.1475-3588.2007.00461.x> PMID: 32847173
47. Nelson EC, Hanna GL, Hudziak JJ, Botteron KN, Heath AC, Todd RD. Obsessive-compulsive scale of the child behavior checklist: specificity, sensitivity, and predictive power. *Pediatrics*. 2001; 108:e14–e14. <https://doi.org/10.1542/peds.108.1.e14> PMID: 11433093
48. Thissen DE, Wainer HE. Lawrence Erlbaum Associates Publishers; 2001.
49. Birnbaum AL. Some latent trait models and their use in inferring an examinee’s ability. *Statistical theories of mental test scores*. 1968;
50. Pirinen T, Kolho KL, Simola P, Ashorn M, Aronen E. Parent–adolescent agreement on psychosocial symptoms and somatic complaints among adolescents with inflammatory bowel disease. *Acta Paediatrica*. 2012; 101:433–437. <https://doi.org/10.1111/j.1651-2227.2011.02541.x> PMID: 22122226
51. Begovac I, Rudan V, Skočić M, Filipović O, Szivovicza L. Comparison of self-reported and parent-reported emotional and behavioral problems in adolescents from Croatia. *Collegium antropologicum*. 2004; 28:393–401. PMID: 15636099

52. Van Der Ende J, Verhulst FC. Informant, gender and age differences in ratings of adolescent problem behaviour. *European Child & Adolescent Psychiatry*. 2005; 14:117–126. <https://doi.org/10.1007/s00787-005-0438-y> PMID: 15959657
53. McCullagh P, Nelder JA. *Generalized linear models*. Routledge; 2019.
54. Angold A, Weissman MM, John K, Merikancas KR, Prusoff BA, Wickramaratne P, et al. Parent and child reports of depressive symptoms in children at low and high risk of depression. *Journal of Child Psychology and Psychiatry*. 1987; 28(6):901–915. <https://doi.org/10.1111/j.1469-7610.1987.tb00678.x> PMID: 3436996