Validation of Strengths and Difficulties Questionnaire (SDQ) for young children in Hong Kong

Reference Guide

Contributing Authors:

Terry Tin-Yau Wong[1], Jacqueline Wai-Yan Tang[1,2], Pokky Poi-Ki Choi[1,2] and Terence Siu-Chung Leung[1]

Reference Guide Preparation Assistance:

Wong Tsz-Yan Annie[2]

[1]Department of Psychology, University of Hong Kong, Pokfulam, Hong Kong SAR, China; [2]Caritas Rehabilitation Service, Hong Kong, Hong Kong SAR, China

Acknowledgement

We would like to express our sincere gratitude to Caritas Rehabilitation Service, especially the colleagues in the Caritas Tier One On-site Pre-school Rehabilitation Services, for their invaluable support in making this validation study of the Strengths and Difficulties Questionnaire (SDQ) in young children in Hong Kong possible. We extend our heartfelt thanks to all the parents and teachers who generously dedicated their time and effort to participate in this research project. Their willingness to complete the questionnaires and share their insights about the children's behavioral and emotional development has been instrumental to the success of this study. Without their meaningful contributions, this important work on establishing culturally appropriate screening tools for Hong Kong's young population would not have been achievable. We hope that the findings from this research will ultimately benefit the broader community of children, families, and professionals working in early childhood development and mental health services in Hong Kong.

The complete findings of this validation study have been published, offering detailed statistical analyses and practical clinical implications for professionals working with preschool children in Hong Kong. For the full publication, please refer to Wong, T. T., Tang, J. W., Choi, P. P., & Leung, T. S. (2025). Identification of preschoolers with special educational needs: Comparing the discriminative validity of the Strengths and Difficulties Questionnaire (SDQ) and the Achenbach System of Empirically Based Assessment (ASEBA) across different informants in Hong Kong. *Frontiers in Psychology, 16*, 1623690. doi:10.3389/fpsyg.2025.1623690

The 26 schools which participated in the research study (in alphabetical order):

Annunciation Catholic Kindergarten
Baptist Oi Kwan Social Service Pui Yan Pre-Primary School
Baptist Pui Li School
Caritas Lok Hing Child Care Centre
Caritas Zonta Club of Hong Kong Nursery School
CCWF King Shing Kindergarten
ECFB Creativity Kindergarten
Immaculate Heart of Mary Kindergarten
Jonathan Innovative English Kindergarten
Kwun Tong Baptist Church Choi Ming Kindergarten
Lok Sin Tong Cheung Yip Mou Ching Kindergarten
Melody Anglo-Chinese Kindergarten (Prime View Garden Branch)
Oi Kwan Rd Baptist Church Lui Kwok Pat Fong Kindergarten
San Po Kong Rhenish Nursery
SKH Crown of Thorns Church Sadick Kindergarten
SKH Good Shepherd Kindergarten
St Andrew's Catholic Kindergarten
St Jerome's Catholic Kindergarten
St Matthias' Church Nursery School
St Peter's Catholic Kindergarten
St Rose of Lima's Kindergarten
Star of the Sea Catholic Kindergarten
Stewards Pooi Chun Kindergarten
Stewards Pooi Yan Kindergarten
Tai Po Catholic Kindergarten
Tsuen Wan Our Lady Kindergarten

Contents

Abstract

Early identification of behavioral and emotional problems in preschoolers, is crucial for providing effective interventions and improving long-term developmental outcomes. The Strengths and Difficulties Questionnaire (SDQ) and the Achenbach System of Empirically Based Assessment (ASEBA) are the most commonly used instruments for identifying children with behavioral and emotional problems. This study aimed to compare the screening efficiency and discriminative validity of the parent and teacher versions of the SDQ and the ASEBA system in identifying preschoolers with special educational needs in Hong Kong.

A community sample (n=312) and a clinical sample (n=79) of children aged 3 to 5 were recruited. Parents and teachers completed the relevant forms: the SDQ-P, SDQ-T, Child Behavior Checklist for Ages 1.5-5 (CBCL 1½-5), and Caregiver-Teacher Report Form (C-TRF). The instruments demonstrated satisfactory internal consistency and test-retest reliability. CBCL exhibited higher internal consistency (Cronbach's α>0.95) compared to the SDQ ($0.70 < \alpha < 0.85$). Interrater reliability between parent and teacher ratings was moderate ($0.26 < rs < 0.36$).

Comparison of discriminative validity showed that teachers' reports were generally more accurate than parents' reports in differentiating the clinical sample from the community sample. Specifically, the SDQ-T yielded the most consistent discriminative validity across all ages (3 to 5), with AUCs consistently above 0.70. Raw scores equal to or above 13, 14, and 11 on the SDQ-T Total Difficulties Score were recommended for ages 3, 4, and 5, respectively as the cutoff values for identifying preschoolers with potential behavioral and emotional problems. These cutoff scores achieve high sensitivity (approx. 0.70) for screening purposes.

This manual serves as a reference guide for psychologists, pediatricians, psychiatrists and allied healthcare professionals in the effective use of these questionnaires when working with preschoolers.

Chapter 1 Introduction

1.1 Purpose of the Study

Behavioral and emotional problems observed during early childhood can have a significant impact on children's long-term development. Young children with such difficulties often experience poorer academic performance (Washbrook et al., 2013) and a higher likelihood of being diagnosed with mental disorders in adolescence (Nielsen et al., 2019). As Educational Psychologists and allied health professionals, it is crucial to differentiate between typical developmental behaviors and those that may signal underlying neurological or psychiatric concerns, such as attention deficit /hyperactivity disorder (ADHD) or autism spectrum disorder (ASD).

To facilitate the effective and early identification of preschoolers who may require further assessment and intervention, the present study aimed to compare the screening efficiency and discriminative validity of two widely used instruments: the Strengths and Difficulties Questionnaire (SDQ) and the Achenbach System of Empirically Based Assessment (ASEBA). The study specifically examined the performance of these tools across different informants (parents vs. teachers) within the Hong Kong context.

Participants, aged 3 to 5 years, were recruited from local preschools to form a representative community sample (n=312) and from On-site Preschool Rehabilitation Services to form a clinical sample (n=79). The study was a collaboration between the Department of Psychology at The University of Hong Kong and Caritas Rehabilitation Service.

1.2 The Instruments

The Strengths and Difficulties Questionnaire (SDQ) and the Achenbach System of Empirically Based Assessment (ASEBA) are two primary, well-established screening instruments utilized internationally for the assessment of child psychopathology. Both are widely-used, informant-rated measures that rely on parents’ and teachers’ reports to capture behavior across different settings. The SDQ is freely available for use in research and clinical work by child psychiatrists, pediatricians and psychologists.

1.2.1 The SDQ

The SDQ is a brief screening questionnaire developed by Goodman (1997) to identify children with mental health issues and special needs. Informants (i.e., parents or teachers) were asked to rate the child on these 25 items using a 3-point Likert scale. Items can be summarized into 5 subscales, namely emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behaviors, with an equal number of items within each domain. The total difficulties score is calculated as the sum of the first four subscales, excluding prosocial behaviors.

The Chinese versions of the SDQs, translated by the Chinese University of Hong Kong, were downloaded directly from the SDQ official website (https://www.sdqinfo.org/) and used in the current study. Two versions of the SDQs were used: the 2–4-year-olds version was used for children aged 3; while the 4–17-year-olds version was used for children aged 4 and 5.

The specific Chinese language versions of the SDQ utilized for this research study are: The SDQ (Chinese Version) for parents or educators of 2-4 year olds and The SDQ (Chinese Version) for parents or teacher of 4-17 year olds. The questionnaires were downloaded directly from the SDQ official website (https://www.sdqinfo.org/).

1.2.2 The CBCL/1½-5 and C-TRF

The Chinese version of CBCL/1½-5 and C-TRF (Leung et al., 2006), two sets of questionnaires in the Achenbach System of Empirically Based Assessment (Achenbach and Rescorla, 2004), were completed by parents and teachers of the participants. CBCL/1½-5 and C-TRF are sets of comprehensive questionnaires tapping various areas of psychopathologies and mental health issues. There are over 100 items within each questionnaire, each of them is rated on a 3-point Likert scale. These items were categorized into six domains in the C-TRF (i.e., emotionally reactive, anxious/depressed, somatic complaints, withdrawn, attention problems, and aggressive behaviors), which were further summarized as internalizing problems (covering the first four domains), externalizing problems (covering the last two domains), and total problems (covering all six domains). The CBCL/1½-5 also included a sleep problem domain, which was not included in either the internalizing or externalizing problems scores but included in the total problem scores.

1.3 Research Study

1.3.1 Sampling Procedure

A. Community Sample

The community sample consists of a total of 312 preschoolers aged 3.0 to 5.11 from 16 preschools and kindergartens in Hong Kong. The participants were recruited through a stratified random sampling procedure, which resulted in a community sample that is representative of the preschool population in Hong Kong in terms of geographical locations and household income by district (see Table 1). The community sample is evenly distributed in terms of age and gender (see Table 2).

Table 1 The geographic location of the community sample in relation to the preschool population in Hong Kong.

Geographic location	Sample Size n (%)	Preschool population in Hong Kong n (%)	Household income by district		
			High	Medium	Low
Hong Kong Island	51 (16%)	26,908 (16%)	51	/	/
Kowloon	99 (32%)	54,561 (33%)	/	54	45
New Territories	162 (52%)	83,466 (51%)	55	56	51
Total	312 (100%)	164,935 (100%)	106	110	96

B. Clinic Sample

A total of 79 preschoolers from kindergartens/kindergarten-cum-child care centres participating in the On-site Preschool Rehabilitation Services in Hong Kong were recruited to comprise the clinical sample. These kindergartens/ kindergarten-cum-child care centres provide preschool rehabilitation services to children with special needs, and only students with diagnoses (e.g., global developmental delay, autism spectrum disorder, attention deficit/hyperactivity disorder, etc.) by pediatricians or psychologists are entitled to these services. The age and gender distributions of the clinical sample are listed in Table 2.

Table 2 The number of participants by age and gender

Age	Community sample			Clinical sample		
	Male	Female	Total	Male	Female	Total
3;0-3;11	50	45	95	15	10	25
4;0-4;11	57	56	113	13	10	23
5;0-5;11	55	49	104	21	10	31
Total	162	150	312	49	30	79

1.3.2 Data Collection

Ethics approval of the current project was obtained from the Departmental Research Ethics Committee of Department of Psychology, he University of Hong Kong. Participating schools and centres helped distribute and collect parental consent from participants' parents. Only participants with parental consent were included in the study. For each participant, two sets of questionnaires were given to their teachers (SDQ-T and C-TRF) and two sets to their parents (SDQ-P and CBCL/1½-5). Questionnaires were distributed in the second semester so that the teachers should have known the children for at least 6 months.

1.3.3 Retest

Within the community sample, a convenient sub-sample of 55 participants was invited for retesting, and their parents and teachers completed the questionnaires again within 1–4 weeks after the initial completion of the questionnaire.

Chapter 2 Statistical Properties

2.1 The SDQ (Chinese Version): Age Effect, Gender Effect, Reliability and Validity

2.1.1 Age and Gender Effect

The total difficulties scores of SDQ-P and SDQ-T, as well as the total problem scores of CBCL/1½-5 and C-TRF, of the community sample participants were analysed using two-way ANOVAs, with age and gender being the independent variables.

A. Age and Gender Effect

The effects of age and gender were only observed in teacher-reported rating scales. A significant main effect of age was observed in SDQ-T only [F (2,262) = 3.99, p = 0.02, η_p^2 = 0.030]. Post-hoc analysis with Bonferroni adjustment revealed a significantly lower total problem difficulties score in SDQ-T in 5-year-olds (M = 8.92, SD = 5.23) than 4-year-olds (M = 11.10, SD = 6.27; p = 0.029). Girls scored lower than boys in both SDQ-T [F (1,262) = 14.04, p < 0.001, η_p^2 = 0.051] and C-TRF [F (1,262) = 6.02, p = 0.015, η_p^2 = 0.022].

Table 3a Age and Gender Effect on the SDQ-P (Chinese Version)

Total *Mean* (*SD*)	Age	Male *Mean* (*SD*)	Female *Mean* (*SD*)	Effect size of age (η_p^2)	Effect size of gender (η_p^2)
12.51 (5.58)	3	13.42 (4.41)	12.85 (4.88)	0.009	0.012
	4	12.98 (4.96)	11.91 (6.91)		
	5	12.87 (6.32)	10.82 (5.36)		

Table 3b Age and Gender Effect on the SDQ-T (Chinese Version)

Total *Mean* (*SD*)	Age	Male *Mean* (*SD*)	Female *Mean* (*SD*)	Effect size of age (η_p^2)	Effect size of gender (η_p^2)
10.28 (5.83)	3	11.91 (5.49)	9.44 (5.75)	0.030*	0.051***
	4	12.65 (5.90)	9.49 (6.28)		
	5	9.89 (5.34)	7.74 (4.92)		

*p < 0.05, ***p < 0.001.

Table 3c Age and Gender Effect on the CBCL/1½-5 (Chinese Version)

Total *Mean* (*SD*)	Age	Male *Mean* (*SD*)	Female *Mean* (*SD*)	Effect size of age (η_p^2)	Effect size of gender (η_p^2)
35.08 (25.53)	3	35.40 (23.89)	37.15 (22.06)	0.003	0.000
	4	35.61 (21.47)	35.60 (28.93)		
	5	35.55 (29.88)	30.67 (26.53)		

Table 3d Age and Gender Effect on the C-TRF (Chinese Version)

Total *Mean* (*SD*)	Age	Male *Mean* (*SD*)	Female *Mean* (*SD*)	Effect size of age (η_p^2)	Effect size of gender (η_p^2)
20.94 (20.63)	3	24.31 (21.97)	18.22 (19.42)	0.017	0.022*
	4	25.88 (19.65)	21.32 (23.78)		
	5	20.98 (21.26)	13.21 (14.16)		

*$p < 0.05$.

B. Age and Gender Interaction

Using two-way ANOVAs, none of the age × gender interaction was found significant.

2.1.2 Reliability

A. Internal Consistency

The internal consistencies of SDQ-P and SDQ-T, as well as those of CBCL/1½-5 and C-TRF, were shown in Table 4. All scales demonstrated satisfactory internal consistency, with Cronbach's α coefficients being greater than 0.7. However, the internal consistencies of CBCL/1½-5 and C-TRF, which were above 0.95, were higher than those of SDQ-P and SDQ-T, which fell between the range of 0.70 to 0.85.

Table 4 Internal Consistency (Cronbach's alpha) of SDQ, CBCL/1½-5 and C-TRF

	Age 3	Age 4-5
SDQ-P	0.710	0.836
SDQ-T	0.825	0.834
CBCL/1½-5	0.959	0.971
C-TRF	0.962	0.962

B. Test-retest Reliability

Test–retest reliabilities, calculated using the intra-class correlations (ICC), were shown in Table 5. Test–retest reliability was satisfactory for all versions (ICCs > 0.80), except for SDQ-T among children aged 4–5 (ICC = 0.67).

Table 5 Test-retest Reliability (intra-class correlations ICC) of SDQ, CBCL/1½-5 and C-TRF

	Age 3	Age 4-5
SDQ-P	0.837	0.820
SDQ-T	0.863	0.670
CBCL/1½-5	0.878	0.888
C-TRF	0.837	0.876

2.1.3 Correlations

Correlations among the total difficulties scores of SDQ-P and SDQ-T, as well as the total problem scores of CBCL/1½-5 and C-TRF, were presented in Table 6. The ratings by the same informants (i.e., SDQ-P with CBCL/1½-5, SDQ-T with C-TRF) correlated strongly with each other (rs > 0.62, ps < 0.01).

The interrater reliability across informants, however, fell only in the moderate range (0.26 < rs < 0.36). The findings suggested a higher level of convergence across instruments than across informants.

Table 6 Correlations among the summary scores

	SDQ-P	SDQ-T	CBCL/1½-5	C-TRF
SDQ-P	-	0.265*	0.626***	0.260*
SDQ-T	0.278***	-	0.196	0.814***
CBCL/1½-5	0.795***	0.195**	-	0.323**
C-TRF	0.351***	0.757***	0.358***	-

Numbers above/below the diagonal represent correlations for age 3/ age 4–5, respectively.
*p < 0.05, **p < 0.01, ***p < 0.001.

2.1.4 Validity

A. Comparison of Group Means

The means and standard deviations of the community sample and the clinical samples on the four summary scores were summarized in Table 7a-d. Due to the main effects of age and gender observed in some of the questionnaires, the use of different forms of SDQ for children aged 3 versus 4–5, as well as the limited number of girls in the clinical sample, the group differences were examined in three separate MANCOVAs, one for each age group, with gender serving as the covariate in these analyses. In general, the parents' ratings were very similar for the community sample and the clinical sample, the only contrasts that were significant were observed among the 4-year-olds [SDQ-P: $F(1,108) = 5.59$, $p = 0.020$, $\eta_p^2 = 0.049$; CBCL/1½-5: $F(1,108) = 13.39$, $p < 0.001$, $\eta_p^2 = 0.138$]. On the other hand, large differences between the community sample and the clinical sample were observed in teachers' ratings, with all the group differences being statistically significant with medium to large effect sizes, F (1,108) s > 5.9, ps < 0.02, $\eta_p^2 \geq 0.05$.

Table 7a. Comparison of the summary scores on the SDQ-P (Chinese Version) between the community sample and the clinical sample

Age	Community sample	Clinical sample	Effect size (η_P^2)
3	13.15 (4.62)	14.47 (4.82)	0.013
4	12.46 (5.99)	16.73 (5.48)	0.049*
5	11.94 (5.97)	12.27 (5.40)	0.000

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 7b Comparison of the summary scores on the SDQ-T (Chinese Version) between the community sample and the clinical sample.

Age	Community sample	Clinical sample	Effect size (η_p^2)
3	10.73 (5.72)	15.79 (5.51)	0.083**
4	11.10 (6.27)	16.60 (3.70)	0.066**
5	8.92 (5.23)	12.80 (5.72)	0.084**

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 7c Comparison of the summary scores on the CBCL/1½-5 (Chinese Version) between the community sample and the clinical sample.

Age	Community sample	Clinical sample	Effect size (η_p^2)
3	36.23 (22.92)	40.05 (20.80)	0.007
4	35.60 (25.27)	67.00 (34.64)	0.138***
5	33.34 (28.35)	34.63 (26.00)	0.000

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 7d Comparison of the summary scores on the C-TRF (Chinese Version) between the community sample and the clinical sample.

Age	Community sample	Clinical sample	Effect size (η_p^2)
3	21.41 (20.90)	45.53 (26.75)	0.101***
4	23.65 (21.77)	48.27 (26.12)	0.110***
5	17.45 (18.69)	29.70 (28.38)	0.050*

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

B. Discriminative Validity

Table 8 shows the results of ROC analyses, which were consistent with the group comparison findings: parent ratings generally did not yield satisfactory area under the curve (AUC) values, except ratings for children aged 4 (SDQ-P AUC = 0.723; CBCL/1½-5 AUC = 0.778). Teachers' ratings, however, yielded higher discriminative validity in terms of differentiating the clinical group from the community group, with five out of six of the AUCs being greater than 0.7. The findings suggested that teachers appear to be better able to differentiate typically developing children from children who may need rehabilitation services. In terms of the comparison across rating scales, SDQ-T appeared to be more consistent in differentiating the clinical sample from the community sample across the age range of 3 to 5, with its AUCs being consistently above 0.70. Given the brevity of the SDQ-T and its consistently good performance in differentiating the clinical samples from the community sample, the SDQ-T total difficulties score was recommended for identifying preschool children who may need rehabilitation services.

Table 8 AUCs of the SDQ, the CBCL/1½-5, and the C-TRF (Chinese Versions)

Age	SDQ-P	SDQ-T	CBCL/1½-5	C-TRF
3	0.576	0.750	0.569	0.759
4	0.723	0.790	0.778	0.771
5	0.525	0.720	0.517	0.650

2.1.5 Cutoff Scores

Based on the score distribution of the SDQ-T total difficulties scores, we explored the sensitivities, specificities, positive predictive values, negative predictive values, and overall classification accuracies across the three age groups by varying the cutoff values (see Chapter 5). While the two cutoff values proposed by Goodman (1997) and Lai et al. (2010) (i.e., 90th and 85th percentiles respectively) resulted in high specificities (SP $\geq$ 0.80), the sensitivities were low (SE $\leq$ 0.47). The cutoff was adjusted downwards to a T score of approximately 54, yielding comparable sensitivities and specificities (Age 3: SE = 0.68, SP = 0.69; Age 4: SE = 0.80, SP = 0.73; Age 5: SE = 0.73, SP = 0.72). Given the purpose of SDQ was to screen children who may need further assessments, sensitivity was valued over specificity, and the cutoff T value of 54 was recommended.

2.1.6 Summary

With parents and teachers as raters, and preschool children as targets, the SDQ-T appeared to perform most consistently in terms of differentiating the clinical sample from the community sample. It showed adequate internal consistency and test–retest reliabilities (except for test–retest reliability of SDQ-T among 4- to 5-year-olds). Its interrater reliability, however, fell only within the moderate range (0.26 < *rs* < 0.36). Convergent validity was confirmed by examining the correlation between the SDQ and the CBCL, showing a strong correlation (*rs* > 0.62, *ps* < 0.01) between ratings completed by the same informants. Given these psychometric properties, the SDQ-T was recommended for the identification of preschoolers with behavioral and emotional problems.

Chapter 3 Administration

3.1 Users

Psychologists, pediatricians, psychiatrists and allied healthcare professionals are recommended to use SDQ, especially teachers' report, or the SDQ-T, to identify preschoolers who may require further assessment for their behavioral and emotional issues.

3.2 Informants

Suitable informants are teachers who have known the child concerned for at least six months. The SDQ-T can be distributed to the teachers, who will have to rate every item in the questionnaires based on their daily observation of the child.

3.3 Target Population

The instruments are suitable for use on preschoolers from 3 to 5 years old, whose parents and teachers have concerns about their behavioral problems, social relationships, as well as their emotional well-being, etc.

3.4 Referral for Medical Diagnosis and Treatment

The SDQ is intended for screening purposes and are not diagnostic tools for specific psychiatric disorders like attention-deficit/hyperactivity disorder (ADHD). However, high scores on the Strengths and Difficulties Questionnaire - Teacher version (SDQ-T) can indicate a need for further evaluation.

Referrals for further assessment by a psychologist or psychiatrist are warranted based on the following:

- Local preschool children who receive a raw score equal to or above 13, 14, and 11 for ages 3, 4, and 5, respectively in SDQ-T are recommended to visit a psychologist or a psychiatrist for further assessments of their developmental and emotional needs.
- Scores Below Cutoff: Even if a child's SDQ-T score does not reach the suggested cut-off of raw scores 13, 14, and 11 for ages 3, 4, and 5 respectively, clinicians should consider other relevant information obtained from other sources, such as interviews, observation, and parental ratings, before making a clinical decision.

3.5 Follow-up

The primary purpose of screening with the Strengths and Difficulties Questionnaire (SDQ) is to identify preschoolers who require further professional evaluation; therefore, the subsequent protocol is focused on rigorous case management, referral, and continuous monitoring. Based on validation studies in the local context, the Teacher version (SDQ-T) is the preferred screening tool for flagging at-risk children due to its superior discriminative validity relative to parent reports. Attaining the cutoff threshold of a T score of 54 or above (equivalent to a raw score of 13, 14, and 11 for ages 3, 4, and 5, respectively) is the key trigger for formal follow-up.

Moreover, since SDQ-T only functions exclusively as a screening instrument and does not constitute a clinical diagnosis, any clinical decision must integrate the screening results with multi-source information, including interviews, observations, and parental ratings, rather than relying on the questionnaire score alone. This reliance on data triangulation is essential, especially given that the study indicated the discriminative validity of the SDQ-T fall only in the satisfactory range (AUC between 0.72 and 0.79).

Chapter 4 Scoring Instructions

4.1 The SDQ (Chinese Version): Scoring Instructions

The Chinese versions of the SDQs adopted in the current study were translated by the Chinese University of Hong Kong and directly downloaded from the SDQ official website. Specifically, the 2–4-year-olds version was used for children aged 3, while the 4-17-year-olds version was used for children aged 4 and 5.

For both the SDQ-P (Chinese Version) and the SDQ-T (Chinese Version), each of the 25 items is scored 0, 1 or 2. These scores correspond to the 3-point Likert scale (typically labeled "Not True," "Somewhat True," and "Certainly True"). The items can be categorized into one of the five subscales:

Subscale	Abbreviation	Item number
Emotional symptoms	E	3, 8, 13, 16, 24
Conduct symptoms	C	5, 7, 12, 18, 22
Hyperactivity symptoms	H	2, 10, 15, 21, 25
Peer problems	P	6, 11, 14, 19, 23
Prosocial behaviors	PS	1, 4, 9, 17, 20

The subscale score is the sum of the 5 items in that subscale. The subscale scores can be transferred to the Record Form in Chapter 5. The Total Difficulties Score is the sum of the scores on all problem scales, i.e., emotional symptoms, conduct symptoms, hyperactivity symptoms, and peer problems. In other words, the sum of scores on all items except those from the subscale prosocial behaviors gives the Total Difficulties Score. The prosocial behaviors subscale score, which is the strengths part of the SDQ, has not been analyzed in the present study.

The sensitivities (SE), specificities (SP), positive predictive value (PPV) and negative predictive value (NPV) for various cutoff values of the Total Difficulties Score are presented in Chapter 5 (5.1).

Chapter 5 Statistical Cutoff Scores

5.1 Statistical Cutoff Scores for the SDQ-T (Chinese Version)

5.1.1 The SDQ-T (Chinese Version) at age 3

	Score	T score	SE	SP	PPV	NPV	Overall acc.
90th Percentile	≥19	64	0.32	0.87	0.35	0.85	0.77
85th Percentile	≥16	59	0.47	0.85	0.41	0.88	0.78
Suggested	≥13	54	0.68	0.69	0.33	0.91	0.69

5.1.2 The SDQ-T (Chinese Version) at age 4

	Score	T score	SE	SP	PPV	NPV	Overall acc.
90th Percentile	≥20	64	0.33	0.90	0.33	0.90	0.82
85th Percentile	≥18	61	0.47	0.83	0.30	0.91	0.78
Suggested	≥14	55	0.80	0.73	0.32	0.96	0.74

5.1.3 The SDQ-T (Chinese Version) at age 5

	Score	T score	SE	SP	PPV	NPV	Overall acc.
90th Percentile	≥17	65	0.20	0.90	0.40	0.76	0.72
85th Percentile	≥13	58	0.43	0.81	0.45	0.80	0.72
Suggested	≥11	54	0.73	0.72	0.48	0.89	0.72

Note: SE = sensitivity, SP = specificity

PPV = positive predictive value; NPV = negative predictive value

References

Achenbach, T. M., and Rescorla, L. A. (2004) The Achenbach System of Empirically Based Assessment (ASEBA) for Ages 1.5 to 18 Years Available online at: https://www.ebsco.com/terms-of-use

Carneiro, A., Soares, I., Rescorla, L., and Dias, P. (2021). Meta-analysis on parent– teacher agreement on preschoolers' emotional and behavioural problems. *Child Psychiatry Hum. Dev.* 52, 609–618. doi: 10.1007/s10578-020-01044-y

Dang, H. M., Nguyen, H., and Weiss, B. (2017). Incremental validity of the child behavior checklist (CBCL) and the strengths and difficulties questionnaire (SDQ) in Vietnam. Asian J. *Psychiatr*. 29, 96–100. doi: 10.1016/j.ajp.2017.04.023

Du, Y., Kou, J., and Coghill, D. (2008). The validity, reliability and normative scores of the parent, teacher and self report versions of the strengths and difficulties questionnaire in China. *Child Adolesc. Psychiatry Ment. Health* 2. doi: 10.1186/1753-2000-2-8

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *J Child Psychol Psychiatry* 38, 581–586. doi: 10.1111/j.1469-7610.1997.tb01545.x

Goodman, A., Heiervang, E., Fleitlich-Bilyk, B., Alyahri, A., Patel, V., Mullick, M. S. I., et al. (2012). Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. *Soc. Psychiatry Psychiatr. Epidemiol*. 47, 1321–1331. doi: 10.1007/s00127-011-0440-2

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., et al. (2016). A systematic review of evidence for the psychometric properties of the strengths and difficulties questionnaire. *Int. J. Behav. Dev*. 40, 64–75. doi: 10.1177/0165025415570647

Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., et al. (2000). Comparing the German versions of the strengths and difficulties questionnaire (SDQ-Deu) and the child behavior checklist. *Eur. Child Adolesc. Psychiatry* 9, 271–276. doi: 10.1007/s007870070030

Koepp, A. E., Watts, T. W., Gershoff, E. T., Ahmed, S. F., Davis-Kean, P., Duncan, G. J., et al. (2023). Attention and behavior problems in childhood predict adult financial status, health, and criminal activity: a conceptual replication and extension of Moffitt et al. (2011)

using cohorts from the United States and the United Kingdom. *Dev. Psychol*. 59, 1389–1406. doi: 10.1037/dev0001533

Lai, K. Y. C., Luk, E. S. L., Leung, P. W. L., Wong, A. S. Y., Law, L., and Ho, K. (2010). Validation of the Chinese version of the strengths and difficulties questionnaire in Hong Kong. *Soc. Psychiatry Psychiatr. Epidemiol*. 45, 1179–1186. doi: 10.1007/s00127-009-0152-z

Leung, P. W. L., Kwong, S. L., Tang, C. P., Ho, T. P., Hung, S. F., Lee, C. C., et al. (2006). Test-retest reliability and criterion validity of the Chinese version of CBCL, TRF, and YSR. J. *Child Psychol. Psychiatry Allied Discip*. 47, 970–973. doi: 10.1111/j.1469-7610.2005.01570.x

Mansolf, M., Blackwell, C. K., Cummings, P., Choi, S., and Cella, D. (2022). Linking the child behavior checklist to the strengths and difficulties questionnaire. *Psychol. Assess*. 34, 233–246. doi: 10.1037/pas0001083

Mulraney, M., Arrondo, G., Musullulu, H., Iturmendi-Sabater, I., Cortese, S., Westwood, S. J., et al. (2022). Systematic review and Meta-analysis: screening tools for attention-deficit/hyperactivity disorder in children and adolescents. *J. Am. Acad. Child Adolesc. Psychiatry* 61, 982–996. doi: 10.1016/j.jaac.2021.11.031

Nielsen, L. G., Rimvall, M. K., Clemmensen, L., Munkholm, A., Elberling, H., Olsen, E. M., et al. (2019). The predictive validity of the strengths and difficulties questionnaire in preschool age to identify mental disorders in preadolescence. *PLoS One* 14. doi: 10.1371/journal.pone.0217707

Rescorla, L. A., Bochicchio, L., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., et al. (2014). Parent-teacher agreement on children's problems in 21 societies. *J. Clin. Child Adolesc. Psychol*. 43, 627–642. doi: 10.1080/15374416.2014.900719

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., and Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clin. Child. Fam. Psychol. Rev.* 13, 254–274. doi: 10.1007/s10567-010-0071-2

Warnick, E. M., Bracken, M. B., and Kasl, S. (2008). Screening efficiency of the child behavior checklist and strengths and difficulties questionnaire: a systematic review. *Child Adolesc. Mental Health* 13, 140–147. doi: 10.1111/j.1475-3588.2007.00461.x

Washbrook, E., Propper, C., and Sayal, K. (2013). Pre-school hyperactivity/attention problems and educational outcomes in adolescence: prospective longitudinal study. *Br. J. Psychiatry* 203, 265–271. doi: 10.1192/bjp.bp.112.123562

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to *ROC. J. Pediatr. Psychol*. 39, 204–221. doi: 10.1093/jpepsy/jst062