
Title	Validity studies of a parent-completed social-emotional measure in a representative sample in China
Author(s)	Huichao Xie, Nicolette Waschl, Xiaoyan Bian, Ruoshui Wang, Chieh-Yu Chen, Luis Anunciação, Zhen Chai, Wei Song and Yan Li

Copyright © 2021 Taylor & Francis

This is an Accepted Manuscript of an article published by Taylor & Francis in *Applied Developmental Science*, on 23/09/2021, available online:

<https://doi.org/10.1080/10888691.2021.1977642>



Validity Studies of a Parent-Completed Social-Emotional Measure in a Representative Sample in China

Journal:	<i>Applied Developmental Science</i>
Manuscript ID	HADS-2021-0017.R2
Manuscript Type:	Empirical Article
Keywords:	social-emotional, factor analysis, reliability, validity, item response theory
Abstract:	<p>In China, more than 90% of individuals in need are not receiving mental health services, partially because of the scarcity of valid and reliable developmental tools. This project aimed to adapt and validate a parent-completed screening tool, the Ages & Stages Questionnaires: Social-Emotional, Second Edition (ASQ:SE-2), to fill in this gap. First, a national representative sample of 2,830 children was accessed to establish cutoff scores. Results from a confirmatory multidimensionality item response theory analysis supported a two-factor structure with this sample. Evidence for item response theory reliabilities and internal consistency were also examined. The second study compared the Chinese ASQ:SE-2 with three convergent measures in a regional sample. Chinese ASQ:SE-2 total scores significantly correlated with most of the domain and total scores on the convergent measures. The classification agreement achieved a maximum of 85%. This project supports the use of the ASQ:SE-2 in the Chinese population, enhancing its clinical utility.</p>

SCHOLARONE™
Manuscripts

VALIDITY OF ASQ:SE-2-C

Tables To Be Inserted (964 Words)

Table 1

Demographic Characteristics of the National Normative Sample (N = 2,830)

	<i>n</i>	%	National census ^a	<i>p</i> value
Region				
Northeast China	242	8.55	7.91 ^{n.s.}	.207
Central & South China	793	28.02	28.26 ^{n.s.}	.777
East China	809	28.59	29.44 ^{n.s.}	.321
North China	374	13.22	12.62 ^{n.s.}	.337
Northwest China	198	7.00	7.31 ^{n.s.}	.526
Southwest China	414	14.63	14.47 ^{n.s.}	.809
Family Registration Status				
Rural	1,398	49.40	42.65 ***	<.0001
Urban	1,432	50.60	57.35 ***	<.0001
Mother's Education				
Junior high and below	1,415	50.00	72.18 ***	<.0001
High school	621	21.94	15.23 ***	<.0001
College and above	789	27.88	12.50 ***	<.0001
Child's Gender				
Male	1,493	52.76	54.33 ^{n.s.}	.094
Female	1,337	47.24	45.67 ^{n.s.}	.094
Child's Ethnic Group				
Han	2,571	90.85	88.67 ***	<.0003
Others	259	9.15	11.33 ***	<.0003

^a National census data retrieved from the China National Bureau of Statistics (2017).

*** $p < .001$.

VALIDITY OF ASQ:SE-2-C

Table 2

Descriptive Statistics of Scores of the National Normative Sample by Interval Including

Those Children Identified as At-Risk Using the Chinese Cutoffs (N = 2,830)

ASQ:SE-2-C Interval	n	Mean	SD	Median	Range	Children Identified as At-Risk	
						n	%
2-month	309	17.65	14.03	15	0–65	49	15.86
6-month	323	19.44	18.90	15	0–120	45	13.93
12-month	310	24.42	24.83	15	0–160	40	12.90
18-month	316	34.06	31.62	25	0–155	31	9.81
24-month	313	34.66	28.42	25	0–180	34	10.86
30-month	315	44.97	35.98	40	0–235	31	9.84
36-month	315	46.78	36.54	40	0–160	38	12.06
48-month	315	48.22	39.16	35	0–230	35	11.11
60-month	314	49.12	36.03	40	0–175	31	9.87
Total	2,830	17.65	14.03	N.A.	0–235	334	11.80

Note: SD = standard deviation; N.A. = not applicable.

VALIDITY OF ASQ:SE-2-C

Table 3

Model Fit Statistics of the Two-Factor Structure in the Chinese National Sample (N = 2,830)

ASQ:SE-2-C Interval	M2	df	<i>p</i>	CFI	TLI	RMSEA	SRMR	<i>r</i>	r_{xx} Emtnl	r_{xx} Social
2-month	154.97	74	< .001	.76	.71	.06	.09	.64	.57	.54
6-month	264.31	189	< .001	.97	.96	.04	.09	.63	.66	.58
12-month	410.51	272	< .001	.97	.96	.04	.09	.74	.69	.70
18-month	536.59	375	< .001	.97	.96	.04	.08	.60	.77	.72
24-month	570.27	374	< .001	.93	.93	.04	.08	.63	.75	.68
30-month	804.33	431	< .001	.91	.91	.05	.09	.84	.80	.77
36-month	1206.47	493	< .001	.88	.87	.07	.11	.80	.82	.79
48-month	981.97	524	< .001	.94	.93	.05	.08	.92	.80	.79
60-month	1031.54	525	< .001	.88	.88	.06	.09	.76	.80	.73

Note: CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square error; *r* = the correlation coefficient between the two factors; r_{xx} Social and r_{xx} Emtnl = IRT reliabilities for the Social and Emotional competence factors, respectively.

VALIDITY OF ASQ:SE-2-C

Table 4

Demographic Characteristics of the Convergent Validity Sample

Demographic Characteristics	
Child's gender	
Male	417 (57.20%)
Female	312 (42.80%)
Child's ethnicity	
Han	704 (98.60%)
Non-Han	10 (1.40%)
Whether the child has been diagnosed with some disabilities	
No	651 (94.90%)
Yes	35 (5.10%)
Person who completed the questionnaires	
Parents	706 (96.71%)
Grandparents and other family members	10 (1.37%)
Non-family caregivers	14 (1.92%)
Child's mother's highest education	
Less than high school	140 (19.42%)
High school	168 (23.30%)
College and above	413 (57.28%)
Annual family income per capita in Chinese RMB	
0–15,000	260 (38.46%)
15,001–25,000	82 (12.13%)
25,001–40,000	90 (13.31%)
40,001 and above	244 (36.09%)

Note: Missing data (ranging from 0 to 7.40% of the sample) were not included.

VALIDITY OF ASQ:SE-2-C

Table 5
Correlation Coefficients Between the ASQ:SE-2-C and the Convergent Measures

		CITSEA	Exter-	Internal-	Dysregu-	Compe-					
ASQ:SE-2-C		naizing	izing	lating	lation	tence					
18M	Total	.31**	.24*	.44**		-.55**					
n = 82	Social	.14 ^{n.s.}	.01 ^{n.s.}	.21 ^{n.s.}		-.57**					
	Emotion	.30**	.39*	.44**		-.21 ^{n.s.}					
24M	Total	.39**	.35**	.55**		-.69**					
n = 73	Social	.26*	.25*	.42**		-.68**					
	Emotion	.32**	.41**	.47**		-.33**					
30M	Total	.35**	.15 ^{n.s.}	.40**		-.67**					
n = 71	Social	.29*	.04 ^{n.s.}	.22 ^{n.s.}		-.72**					
	Emotion	.33**	.31**	.50**		-.38**					
ASQ:SE-2-C		CBCL 2-3	Disrup-	Aggres-	Somatic	With-	Depre-	Sleep	Exter-	Internalizing	Total
			tive	sive	Com-	drawal	ssion	Problems	nalizing		
					plaints						
36M	Total	.51**	.53**	.46**	.63**	.51**	.37**	.54**	.61**	.61**	

VALIDITY OF ASQ:SE-2-C

<i>n</i> = 97	Social	.46**	.41**	.34**	.59**	.46**	.24*	.44**	.56**	.51**	
	Emotion	.50**	.56**	.43**	.40**	.36**	.36*	.57**	.39**	.55**	
48M	Total	.45**	.46**	.31**	.62**	.45**	.26*	.48**	.57**	.53**	
<i>n</i> = 91	Social	.38**	.33**	.27*	.62**	.44**	.20 ^{n.s.}	.35**	.56**	.44**	
	Emotion	.45**	.54**	.30**	.47**	.39**	.37**	.53**	.46**	.53**	
	CBCL 4–18	With-	Somatic	Anxiety/	Sexual	Schi-	Aggress-	Imma-	Conduct	External-	Total
		drawal	Com-	Depres-	Prob-	zoid	ive	ture/	Problems	izing	
	ASQ:SE-2-C		plaints	sion	lems			Obese	/Hyperactive		
60M	Total	.34**	.25**	.41**	.25**	.31**	.48**	.49**	.29**	.46**	.39**
<i>n</i> = 316	Social	.28**	.18**	.30**	.24**	.25**	.30**	.38**	.20**	.30**	.30**
	Emotion	.28**	.24**	.36**	.19**	.25**	.44**	.42**	.28**	0.43**	.34**

* $p < .05$; ** $p < .01$; n.s. = not significant

Note: 18M = 18-month; 24M = 24-month; 30M = 30-month; 36M = 36-month; 48M = 48-month; 60M = 60-month; ASQ:SE-2-C = Ages & Stages Questionnaires: Social-Emotional, Second Edition in Chinese; CITSEA = Chinese Infant-Toddler Social and Emotional Assessment; CBCL 2-3 = Child Behavior Checklist for two to three-year-old Children; CBCL 4-18 = Child Behavior Checklist for Four- to 18-year-old Children.

VALIDITY OF ASQ:SE-2-C

Table 6

*Indicators for Classification Agreement Between the ASQ:SE-2-C and Corresponding**Convergent Measures by Age Interval*

Age	n	Convergent measure	Overall agreement	Sensitivity	Specificity	Under-IDed.	Over-IDed.
18	82	CITSEA	76%	33%	88%	15%	10%
24	73	CITSEA	82%	67%	90%	11%	7%
30	71	CITSEA	79%	60%	93%	17%	4%
36	97	CBCL 2-3	80%	92%	79%	1%	19%
48	91	CBCL 2-3	78%	73%	79%	4%	18%
60	316	CBCL 4-18	85%	77%	88%	7%	8%

Note: IDed. = Identification.

VALIDITY OF ASQ:SE-2-C

1

Supplementary Tables

Table S1.

Internal Consistencies of the ASQ:SE-2-C in Total by Factor

ASQ:SE-2-C Interval	ASQ:SE-2-C		Social Competence		Emotional Competence	
	Ord. α	No. of Items	Ord. α	No. of Items	Ord. α	No. of Items
2-month	.81	16	-	-	-	-
6-month	.89	22	.89	12	.77	10
12-month	.94	26	.94	16	.80	10
18-month	.93	30	.92	18	.89	12
24-month	.92	30	.91	16	.85	14
30-month	.93	32	.90	15	.88	17
36-month	.93	34	.90	15	.89	19
48-month	.94 ^a	35	.92	15	.89	20
60-month	.93 ^a	35	.86	15	.90	20

Note. ^a Item 6 showed a negligible negative item-total correlation.

VALIDITY OF ASQ:SE-2-C

Table S2

Factor loadings for 2-month interval.

Item	F1	F2
Item 1	.67	
Item 2		.34
Item 3	.52	
Item 4		.81
Item 5		.67
Item 6		.81
Item 7	.48	
Item 8	.58	
Item 9		.59
Item 10	.61	
Item 11	.34	
Item 12		.61
Item 13	.49	
Item 14		.37
Item 15	.36	

VALIDITY OF ASQ:SE-2-C

3

Table S3

Factor loadings for 6-month interval.

Item	F1	F2
Item 1	.47	
Item 2		.83
Item 3		.46
Item 4	.49	
Item 5		.67
Item 6		.65
Item 7		.77
Item 8	.51	
Item 9	.68	
Item 10		.78
Item 11	.55	
Item 12	.45	
Item 13		.83
Item 14	.72	
Item 15		.44
Item 16	.58	
Item 17	.54	
Item 18	.59	
Item 19		.84
Item 20		.81
Item 21		.88
Item 22		.72

VALIDITY OF ASQ:SE-2-C

Table S4

Factor loadings for 12-month interval.

Item	F1	F2
Item 1		.36
Item 2		.65
Item 3		.66
Item 4		.60
Item 5	.76	
Item 6	.53	
Item 7		.62
Item 8		.89
Item 9	.59	
Item 10	.66	
Item 11		.85
Item 12	.33	
Item 13		.52
Item 14	.60	
Item 15	.67	
Item 16		.78
Item 17	.65	
Item 18	.62	
Item 19		.77
Item 20		.85
Item 21	.41	
Item 22		.73
Item 23		.79
Item 24		.73
Item 25		.84
Item 26		.74

VALIDITY OF ASQ:SE-2-C

5

Table S5

Factor loadings for 18-month interval.

Item	F1	F2
Item 1		.56
Item 2	.56	
Item 3		.73
Item 4		.63
Item 5		.80
Item 6		.31
Item 7	.42	
Item 8	.74	
Item 9	.78	
Item 10		.62
Item 11	.65	
Item 12	.73	
Item 13	.72	
Item 14		.51
Item 15	.52	
Item 16		.86
Item 17	.56	
Item 18		.70
Item 19		.83
Item 20		.72
Item 21		.58
Item 22		.46
Item 23	.84	
Item 24		.63
Item 25	.68	
Item 26		.77
Item 27		.83
Item 28		.53
Item 29	.53	
Item 30		.79

VALIDITY OF ASQ:SE-2-C

Table S6

Factor loadings for 24-month interval.

Item	F1	F2
Item 1		.60
Item 2	.23	
Item 3		.60
Item 4		.63
Item 5	.71	
Item 6		.46
Item 7		.10
Item 8	.44	
Item 9	.65	
Item 10		.56
Item 11	.67	
Item 12		.57
Item 13	.60	
Item 14	.30	
Item 15		.77
Item 16	.60	
Item 17	.55	
Item 18		.85
Item 19		.78
Item 20		.60
Item 21	.57	
Item 22		.58
Item 23	.77	
Item 24		.52
Item 25	.62	
Item 26		.73
Item 27		.53
Item 28	.66	
Item 29		.79
Item 30	.63	

VALIDITY OF ASQ:SE-2-C

7

Table S7

Factor loadings for 30-month interval.

Item	F1	F2
Item 1		.78
Item 2		.12
Item 3	.49	
Item 4		.41
Item 5	.80	
Item 6		.44
Item 7	.23	
Item 8	.77	
Item 9	.59	
Item 10	.66	
Item 11	.45	
Item 12	.68	
Item 13		.68
Item 14	.61	
Item 15	.67	
Item 16		.72
Item 17		.85
Item 18	.75	
Item 19		.63
Item 20		.69
Item 21		.75
Item 22	.61	
Item 23		.64
Item 24	.61	
Item 25	.54	
Item 26		.68
Item 27	.51	
Item 28		.75
Item 29		.62
Item 30		.61
Item 31	.57	
Item 32	.60	

VALIDITY OF ASQ:SE-2-C

Table S8

Factor loadings for 36-month interval.

Item	F1	F2
Item 1		.65
Item 2		.32
Item 3		.64
Item 4	.57	
Item 5	.75	
Item 6	.36	
Item 7	.74	
Item 8	.62	
Item 9	.68	
Item 10		.46
Item 11	.55	
Item 12	.59	
Item 13	.68	
Item 14		.67
Item 15	.49	
Item 16	.77	
Item 17		.93
Item 18		.66
Item 19	.62	
Item 20		.74
Item 21	.65	
Item 22	.65	
Item 23		.76
Item 24	.55	
Item 25		.76
Item 26		.73
Item 27		.89
Item 28		.77
Item 29	.60	
Item 30	.58	
Item 31		.78
Item 32		.45
Item 33	.67	
Item 34	.61	

VALIDITY OF ASQ:SE-2-C

9

Table S9

Factor loadings for 48-month interval.

Item	F1	F2
Item 1		.69
Item 2	.39	
Item 3		.64
Item 4	.75	
Item 5		.27
Item 6	.02	
Item 7	.60	
Item 8	.55	
Item 9		.66
Item 10		.50
Item 11	.57	
Item 12		.70
Item 13	.60	
Item 14	.91	
Item 15	.62	
Item 16	.45	
Item 17		.85
Item 18	.76	
Item 19		.83
Item 20	.73	
Item 21		.39
Item 22	.43	
Item 23	.73	
Item 24	.72	
Item 25	.51	
Item 26		.67
Item 27		.73
Item 28		.66
Item 29		.77
Item 30		.60
Item 31	.65	
Item 32	.29	
Item 33	.46	
Item 34	.52	
Item 35		.76

VALIDITY OF ASQ:SE-2-C

Table S10

Factor loadings for 60-month interval.

Item	F1	F2
Item 1		.64
Item 2	.53	
Item 3		.17
Item 4		.51
Item 5	.57	
Item 6	.08	
Item 7	.67	
Item 8	.79	
Item 9	.55	
Item 10		.50
Item 11		.71
Item 12	.71	
Item 13	.55	
Item 14		.60
Item 15	.69	
Item 16	.60	
Item 17	.73	
Item 18		.98
Item 19		.82
Item 20	.61	
Item 21		.22
Item 22	.58	
Item 23	.77	
Item 24	.74	
Item 25	.60	
Item 26		.52
Item 27		.57
Item 28		.71
Item 29		.72
Item 30	.76	
Item 31		.63
Item 32	.43	
Item 33	.46	
Item 34	.76	
Item 35		.69

1
2
3
4
5
6
7
8
9

**Validity Studies of a Parent-Completed Social-Emotional Measure
in a Representative Sample in China**

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

Abstract (150 words)

In China, more than 90% of individuals in need are not receiving mental health services, partially because of the scarcity of valid and reliable developmental tools. This project aimed to adapt and validate a parent-completed screening tool, the Ages & Stages Questionnaires: Social-Emotional, Second Edition (ASQ:SE-2), to fill in this gap. First, a national representative sample of 2,830 children was accessed to establish cutoff scores. Results from a confirmatory multidimensionality item response theory analysis supported a two-factor structure with this sample. Evidence for item response theory reliabilities and internal consistency were also examined. The second study compared the Chinese ASQ:SE-2 with three convergent measures in a regional sample. Chinese ASQ:SE-2 total scores significantly correlated with most of the domain and total scores on the convergent measures. The classification agreement achieved a maximum of 85%. This project supports the use of the ASQ:SE-2 in the Chinese population, enhancing its clinical utility.

39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Key words: social-emotional, factor analysis, reliability, validity, item response theory

(Main text in 10783 Words, including references and tables)

Validity Studies of a Parent-Completed Social-Emotional Measure in a Representative Sample in

China

Introduction

The Significance of Early Childhood Social-Emotional Screening

The development of social-emotional competence is to “form close and secure adult and peer relationships; experience, regulate, and express emotions in socially and culturally appropriate ways; and explore the environment and learn” (Yates et al., 2008, p. 2). Further, this competence is a multidimensional concept and can include dimensions such as self-perception, emotion expression, emotion management, attachment, empathy, perspective taking, and inhibitory control (Cantor, Osher, Berg, Steyer, & Rose, 2019; Denham, Wyatt, Bassett, Echeverria, & Knox, 2009). In general, social-emotional competence can also be defined by two closely connected but distinct dimensions: social and emotional competence (Squires, Bricker, & Twombly, 2015). In this bi-dimensional definition, social competence supports children in having a positive relationship with others (Jones & Bouffard, 2012; Raver & Zigler, 1997), while emotional competence aids them in regulating their emotions to reach goals (Campos, Mumme, Kermoian, & Campos, 1994). Moreover, young children without appropriate social-emotional competence are more likely to engage in challenging behaviors, and these tend to persist over time (Carter, Briggs-Gowan, & Davis, 2004). Thus, early and accurate identification of any possible problems or delays in children’s social-emotional development is essential in preventing further deficits and providing timely and early interventions (Bricker, Macy, Squires, & Marks, 2013). For example, Reid et al. (2020) found that parental reports on children’s social-emotional development at 18 months of age indicated the potential of being diagnosed with autism at three

years old. In this context, the current study aimed to examine the validity of evidence of a parental report measure of young children's social-emotional development in China, where limited early childhood mental health professional resources are faced with increasing needs for services.

The Need for Cost-Effective Social-Emotional Screening Tools in China

China is a country with a population of around 1.4 billion (The World Bank, 2021). Despite its size, there is still a significant gap between the need for the timely identification of problems in young children's mental health and available resources. According to the most updated available estimates for 2015, the number of children aged 0–6 in China was 95.31 million (UNICEF [United Nations International Children's Emergency Fund], 2016). Moreover, a synthesis study indicated that the prevalence of mental health problems in preschool children ranged from 6% to 36% (Chen, Li, Wei, & Zhang, 2015; Ning, He, & Yu, 2017; Ye & Tan, 2015). However, significant and profound gaps were reported in the availability and distribution of mental health resources in China (Que, Lu, & Shi, 2019). The statistics also reveal that more than 90% of individuals in need do not receive affordable mental health services (Liu et al., 2011).

Mental health in the early years of life has gained increasing attention among policy makers and professionals serving young children and their families. Further, an action plan released by the Chinese government (China National Health Commission, 2019) required preschools and child healthcare providers to offer mental health services by the end of 2022. In addition, the recent COVID-19 outbreak has shed more light on the urgent needs of children in relation to mental well-being. In this context, a nationwide survey indicated significant increases in several risk factors and symptoms such as spending long periods of time on digital devices,

1
2
3 irregular routines, sleep problems, hyperactive behaviors, difficulties staying focused, and more
4
5 tantrums (Liang et al., 2020). These conditions increase the importance of exploring cost-
6
7 effective approaches to meet the growing needs in such a resource-limited context.
8
9

10 **Parent-completed Screening Tools as a Possible Solution**

11
12 Existing procedures used to identify behavioral and emotional problems in early
13
14 childhood take the form of parent- or caregiver-report questionnaires and direct observations of
15
16 the child (Bagner, Rodríguez, Blake, Linares, & Carter, 2012). Using parent- or caregiver-report
17
18 questionnaires has been recommended for the reasons listed below. First, involving parents in the
19
20 assessment with their child ensures that parents' extensive knowledge about their child is
21
22 included (Sandall, Hemmeter, Smith, & McLean, 2005). Lack of parental input in the clinical
23
24 assessment of a child's social-emotional development may result in insufficient or inaccurate
25
26 data for decision-making. Second, compared to a traditional, direct testing approach, using
27
28 parent-completed screening tools to inform clinical referral requires fewer professional personnel
29
30 resources and therefore is more convenient, especially for the ongoing monitoring of social-
31
32 emotional development that requires repeated screenings (Drotar, Stancin, & Dworkin, 2008;
33
34 Glascoe & Robertshaw, 2007; Pizur-Barnekow et al., 2010). Third, parent-completed screening
35
36 tools are flexible in their administration and less time consuming for clinical professionals and
37
38 are thus more appropriate for large-scale screening purposes (Squires et al., 2015). A systematic
39
40 review of behavioral and emotional measures in infancy (Bagner et al., 2012; Pontoppidan, Niss,
41
42 Pejtersen, Julian, & Vaever, 2017) reported specific strengths of parent-completed tools,
43
44 including high internal consistency and validity with other measures of related constructs.
45
46
47
48
49
50

51 However, existing early childhood social-emotional measures in China have several
52
53 limitations. First, most Chinese measures, such as the Chinese Infant-Toddler Social and
54
55
56
57
58
59
60

VALIDITY OF ASQ:SE-2-C

5

1
2
3 Emotional Assessment (CITSEA, Wang et al., 2009) and the Chinese Child Behavior Checklist
4 scales (CBCL, Liu, Wu, & Yao, 2003; Xi et al., 1992), focus on problematic behaviors rather
5
6 than the social-emotional competencies of young children. As suggested in the literature
7
8 (Pontoppidan et al., 2017), strengths-focused rather than problem-focused measures may be more
9
10 popular with caregivers. Second, some strengths-focused measures—such as the Devereux Early
11
12 Childhood Assessment for Infants and Toddlers (Powell, Mackrain, & LeBuffe, 2007) and the
13
14 Greenspan Social-Emotional Growth Chart in the Bayley Scales of Infant Development and
15
16 Toddler Development (Bayley, 2006)—cover a shorter age range up to 36 or 42 months, which
17
18 are deemed insufficient in China where early childhood services are expected to cover 1- 72
19
20 months of age (Zhu & Zhang, 2008). Meanwhile, Ages & Stages Questionnaires, Second Edition
21
22 (ASQ:SE-2; Squires et al., 2015) has offered a promising option for its parent-completed feature
23
24 with the inclusion of a majority of strengths-based items, sound studies of its psychometric
25
26 properties, and coverage of a wider age ranging up to 72 months.
27
28
29
30
31

Ages & Stages Questionnaires, Second Edition (ASQ:SE-2)

32
33
34
35 The ASQ:SE-2 is a set of parent-completed screening questionnaires for children aged 1–
36
37 72 months. As a social-emotional screening measure, the ASQ:SE-2 was developed to measure
38
39 whether a child’s social-emotional development is typical or potentially problematic (i.e., “at
40
41 risk”), given their chronological age (Bricker et al., 2013). The ASQ:SE-2 contains a set of nine
42
43 questionnaires, each targeting a specific range of the child’s age (e.g., the 36-month
44
45 questionnaire is for children from 33 months 0 days to 41 months 30 days of age). The cutoff
46
47 scores of the ASQ:SE-2 increase its clinical utility, as they can be used as an index to make
48
49 referrals or follow-up decisions (Squires et al., 2015).
50
51
52

53
54 According to Bagner et al. (2012), parent- and caregiver- completed measures, such as
55
56
57
58
59

VALIDITY OF ASQ:SE-2-C

6

1
2
3 the ASQ:SE, have the most extensive psychometric evidence when compared to clinician-
4 completed and observational measures. In the U.S., where the ASQ:SE-2 was developed, strong
5
6 completed and observational measures. In the U.S., where the ASQ:SE-2 was developed, strong
7
8 psychometric evidence has been established regarding the internal consistency, test-retest
9
10 reliability, and convergent validity of its scores (Squires et al., 2015). For convergent validity,
11
12 the ASQ:SE-2 screening results (i.e., at-risk or typical development) were compared with results
13
14 on several other social-emotional measures, including the Devereux Early Childhood
15
16 Assessment for Infants and Toddlers (Powell et al., 2007), the Infant Toddler Social and
17
18 Emotional Assessment (Carter & Briggs-Gowan, 2006), and the Child Behavior Checklist
19
20 (Achenbach & Rescorla, 2000). In this context, the results of a convergent validity study
21
22 (Squires et al., 2015) indicated acceptable evidence for sensitivity, ranging from 77.8% (2-
23
24 month) to 84.0% (24-month) and specificity, ranging from 76.2% (18-month) to 98.0% (60-
25
26 month). Two separate studies reported the good fit of the two-factor (social competence and
27
28 emotional competence) structure for the ASQ:SE-2 in U.S. (Chen, Squires, & Scalise, 2020b)
29
30 and in Taiwanese samples (Chen, Squires, Chen, Wu, & Xie, 2020a).
31
32
33
34

35
36 Meanwhile, the first edition of the Ages & Stages Questionnaires: Social-Emotional
37
38 (Squires, Bricker, & Twombly, 2002) has been translated and validated for use in China (Bian,
39
40 Xie, Squires, & Chen, 2017), following the translation and backward translation to ensure
41
42 cultural sensitivity (International Test Commission, 2005). Further, another study compared the
43
44 screening results on the Chinese ASQ:SE with several Chinese social-emotional and behavioral
45
46 measures commonly used in clinical practices and reported significant but lower-than-expected
47
48 correlations (Xie, Bian, Chen, Squires, & Lu, 2019).
49
50

51
52 The current study describes an important update of the Chinese ASQ:SE to its second and
53
54 refined version. This updated ASQ:SE-2 includes additional items to elicit parent concerns
55
56
57
58
59
60

VALIDITY OF ASQ:SE-2-C

7

related to autism (Squires et al., 2015), which is particularly important due to the consistently lower identification rate of children with autism in China (Pang et al., 2018). Therefore, the ASQ:SE-2 has great potential to enhance the identification of young children at risk of social-emotional problems in China. However, research has indicated that parental perceptions and reports on social-emotional skills and problems may differ across cultures (e.g., Chung et al., 2012; Matson et al., 2011). As a parent-report measure, the validity of the ASQ:SE-2 in Mainland China remains unknown. To date, the two-factor (social competence and emotional competence) structure of the ASQ:SE, either the first or the second edition, has not been examined on young children in Mainland China. Thus, the current examination was conducted to fill this gap.

Aims of the Current Examination

The current study has three aims. The first is to update the Chinese ASQ:SE to the most recent version, the ASQ:SE-2, by performing a culturally contextualized translation of the new items into Chinese and establishing a national sample and cutoff scores. The second is to examine whether the two-factor structure fits well for the Chinese ASQ:SE-2—which has been demonstrated in studies with the ASQ:SE-2 in the U.S. (Chen et al., 2020b) and Taiwanese contexts (Chen et al., 2020a)—as well as for the ASQ:SE, First Edition, which has been examined in studies across the U.S. (Chen et al., 2016) and Brazil (Anuniação, Squires, Clifford, & Landeira-Fernandez, 2019). Finally, the third aim is to compare the screening results of the adapted ASQ:SE-2 Chinese version (ASQ:SE-2-C) with the results of convergent measures to further examine score validity. As a screening measure, the ASQ:SE-2-C was expected to demonstrate significant correlations with the convergent measures at .70 with the competence domain on the CITSEA and relatively lower correlations with the problem-focused

VALIDITY OF ASQ:SE-2-C

8

CITSEA domains (i.e., internalizing, externalizing, and dysregulation) and measures (i.e., CBCL for 2-3 years of age, as CBCL 2-3; CBCL for 4-18 years of age, as CBCL 4-18). However, a previous study (Xie et al., 2019) reported that the ASQ:SE-C, First Edition, showed statistically significant but lower-than-expected correlations (.26 to .70) with the CITSEA, CBCL 2-3, and CBCL 4-18. In the current study (i.e., Study 2), ASQ:SE-2-C is expected to show similar correlations with these convergent measures.

Two empirical studies were conducted to address these aims. The first sought to establish a national normative sample and test the factor structure of the ASQ:SE-2-C. The second sought to provide validity evidence for ASQ:SE-2-C scores via a convergent analysis of them and other variables using a regional sample of children.

Study 1**Methods*****Participants***

A stratified purposeful sampling method was used to ensure the distribution of the national sample was similar to that of the 2016 national census data (China National Bureau of Statistics, 2017) with regard to geographic region, family registration status (rural or urban), child's gender, and ethnicity. Thus, the Chinese normative sample aimed to recruit at least 300 children in each of the nine ASQ:SE-2-C age intervals, based on the sample sizes of the previous studies of the ASQ:SE, First Edition, in the U.S. (Squires et al., 2002; the sample size ranged from 298 to 471) and China (Bian et al., 2017; the sample size ranged from 305 to 330). Within each of the six national regions across China (Northeastern, Central and Southern, Eastern, Northern, Northwestern, and Southwestern), one city (urban) or county (rural) was randomly selected from those with a medium level gross domestic product (GDP) (i.e., within the 40th to

VALIDITY OF ASQ:SE-2-C

9

60th percentile range) in the region. The recruitment team approached 3,212 potential participants, of which 2,830 (88.11%) completed and submitted questionnaires. Here, the following reasons reported by the staff at the recruitment sites might explain why some questionnaires were not returned: caregivers who were not the parent of the child (e.g., a grandparent or an aunt) were more likely to reject the invitation to participate as parental consent was required; caregivers often responded with “I am in a rush” when asked why they declined to participate. Out of the 2,830 returned questionnaires, only 42 (1.48%) were incomplete with missing values, mostly due to the caregiver and child having to rush to the next appointment before completing the questionnaire. Given the low proportion of missing values, a conservative approach, which is the listwise deletion procedure, was used to handle the missing cases, as recommended by the literature (Dong & Peng, 2013; Graham, 2012).

Measures

The ASQ:SE-2-C, updated from the first version of the Chinese ASQ:SE, was used to collect information about the participating children’s social-emotional development. The first version of the Chinese ASQ:SE was created from the English version using a translation and back-translation procedure to ensure accuracy (International Test Commission, 2017), and two types of adaptations (adding definitions and examples) were made to improve the relevance to the Chinese context (Bian et al., 2017). The 19 new items found in the second edition in English were translated following the same translation procedure mentioned above by three of the authors of this paper to enhance the linguistic equivalence between the original and Chinese versions of the ASQ:SE-2. First, the third author of this paper, a native Chinese speaker who has more than 30 years of practice as a pediatrician in China, translated the 19 new items from English to simplified Chinese. Then the first author, who practiced as an early interventionist in

China before completing her doctoral training in the U.S., and the fourth author, who graduated from a medical school in China and has been accepted to a doctoral program in the U.S., back-translated the new items to English and evaluated the linguistic equivalence to inform revisions and finalize the Chinese translation. As described earlier, the ASQ:SE-2-C is a screening tool consisting of nine questionnaires at 2, 6, 12, 18, 24, 30, 36, 48, and 60 months. Each questionnaire consists of 16 to 36 items and takes about 10–20 minutes for a parent or caregiver to complete (Squires et al., 2015). Each item can be rated with three possible responses. For an item targeting a skill (e.g., “*Does your child seem happy?*” in the 30-, 36-, 48-, and 60-month questionnaires), the points for the three responses are 0 for “often or always,” 5 for “sometimes,” and 10 for “rarely or never.” Meanwhile, for an item targeting a problem behavior, the points are 10 for “often or always,” 5 for “sometimes,” and 0 for “rarely or never.” However, any item with the “concern” box checked by the parent adds an extra 5 points to the total score. The total score on a questionnaire is compared to a cutoff score, and a monitoring range derived from the normative sample is used to determine the screening result for a child: a score higher than the cutoff indicates the need for a referral for further assessment and/or intervention services; a score lower than the cutoff but within the monitoring range indicates a need to monitor the child’s progress and pursue follow-up actions for items of concern; finally, a score lower than the monitoring range indicates that the child’s social-emotional development appears to be typical (Squires et al., 2015).

Procedures

Recruitment and data collection took place in 2017 in maternal and child health clinics in the selected counties of the six regions. Inclusion criteria were children aged from one to 72 months who were registered as residents in the county. Exclusion criteria were children whose

VALIDITY OF ASQ:SE-2-C

11

1
2
3 age did not fall in the ASQ:SE-2 age range during the time of recruitment and those who were
4
5 not residents. Recruitment flyers were given to caregivers bringing their children to the clinics
6
7 for well-child checkups or other healthcare services at a ratio of one in every four eligible
8
9 children. Once consent was given, the caregiver and the child were placed in a quiet room in the
10
11 clinic to complete the ASQ:SE-2-C. Staff at each data collection clinic were trained to provide
12
13 support as needed, such as arranging referrals for diagnosis, providing parent–child activity
14
15 suggestions, and supplying information about local resources for early intervention. Participants
16
17 who returned the questionnaire received a one-page, age-appropriate recommendation for parent-
18
19 child activities at home (Squires et al., 2015) as a token of appreciation. No monetary
20
21 compensation was provided. Ethical approval for the two studies was granted by the Shanghai
22
23 Normal University Institutional Review Board (IRB) (Protocol ID 392020).
24
25
26
27

Data Analyses

28
29
30
31 To evaluate the factor structure of the ASQ:SE-2-C across all age intervals (two to 60
32
33 months), a confirmatory multidimensional item response theory (MIRT) analysis was used. All
34
35 the models tested were based on the social and emotional factors previously reported in Chen et
36
37 al. (2020b). As recommended by the literature, the models tested did not include the last item in
38
39 each age interval (“*Has anyone shared concerns about your baby’s [or child’s] behaviors?*
40
41 *Please explain.*”) because it refers to general concerns rather than a specific social-emotional
42
43 skill or problem. Therefore, it is not suitable for analysis within an item response theory (IRT)
44
45 framework. The removal of the last scoring item resulted in 15–35 items for each interval.
46
47
48

49
50 The IRT methodology was selected to examine the factor structure and reliability of the
51
52 ASQ:SE-2-C as it represents a qualitative improvement from the Classical Test Theory (CTT).
53
54 Moreover, IRT models the relation between true scores and latent variables (Thomas, 2011) and
55
56
57
58
59
60

1 provides the ability to evaluate how each item contributes to the development of social-
2
3 emotional skills. Furthermore, in modeling categorical item data, a choice can be made between
4
5 limited information methods arising from the factor analysis (CTT) tradition and full information
6
7 methods arising from the IRT tradition (Forero & Maydeu-Olivares, 2009). Although studies
8
9 have found that these two methods generally have similar performances, full information
10
11 methods may be advantageous for smaller samples with less than 500 observations (Forero &
12
13 Maydeu-Olivares, 2009), as in the present case where the age interval sample sizes ranged from
14
15 309 to 323. Additionally, full information IRT estimation allows for direct tests of the
16
17 approximate model fit (Maydeu-Olivares & Joe, 2006, 2014), while limited information
18
19 confirmatory factor analysis estimation typically relies on sample statistics, such as thresholds
20
21 and polychoric correlations, in determining model fit, which may be problematic under certain
22
23 conditions (Forero & Maydeu-Olivares, 2009).
24
25
26
27
28
29

30
31 Meanwhile, Samejima's (1969) graded response model (GRM) was estimated using
32
33 marginal maximum likelihood estimation via the expectation-maximization algorithm. The GRM
34
35 is a generalization of the two-parameter logistic IRT model, which allows the discrimination (a ,
36
37 slope) and location (b , threshold) parameters to vary across all items. Moreover, model fit was
38
39 evaluated using fit indices based on the M2 statistic (Cai & Hansen, 2013; Maydeu-Olivares &
40
41 Joe, 2006, 2014). In accordance with the best practices (Brown, 2006; Hu & Bentler, 1999), the
42
43 following standard recommendations for fit indices were used: $\geq .95$ for the comparative fit
44
45 index (CFI) and Tucker-Lewis index (TLI); $\leq .08$ for the standardized root mean residual
46
47 (SRMR); and $\leq .06$ for the root mean square error of approximation (RMSEA).
48
49
50

51 Ordinal alpha values were calculated from the polychoric correlation matrices of the item
52
53 scores (Gadermann, Guhn, & Zumbo, 2012). Reliability was also calculated under an IRT
54
55
56
57
58
59

framework for the social and emotional factors (Raju, Price, Oshima, & Nering, 2007). All psychometric analyses in Study 1 were performed in R Version 3.6.1 (R Core Team, 2019), with the “mirt” (Chalmers, 2012) and “psych” (Revelle, 2019) packages.

Results

Establishing a National Normative Sample and Cutoff Scores

A total of 2,830 children and their caregivers participated in the Chinese national normative sample. Among the responding caregivers, 2,374 (83.89%) were either the father, the mother, or both parents working together to complete the ASQ:SE-2-C, while 387 (13.67%) were grandparents. In cases when more than one caregiver (e.g., both parents, or one parent and a grandparent) was involved in completing the ASQ:SE-2-C, only one form was submitted per child. Moreover, most of the participating families with “rural” ($n = 1,000$, 71.53%) and “urban” registration statuses ($n = 892$, 62.30%) reported annual family incomes per capita that were lower than the 40th percentile among families with the same registration status in China ($n = 7,828$ RMB for rural families, 23,055 RMB for urban families; China National Bureau of Statistics, 2017). Table 1 shows the sample distribution by geographic region, rural and urban type of family registration, child’s mother’s highest education, and child’s gender and ethnicity.

[insert Table 1 here.]

Three methods for establishing ASQ:SE-2 cutoff scores have been reported. The original ASQ:SE-2 established cutoffs using the receiver operating characteristics (ROC) curve, which was not applicable to the Chinese version due to the lack of concurrent measures in Chinese for infants. In this context, the developers of the original ASQ:SE-2 had suggested two alternative methods for establishing cutoffs when the ROC method is not applicable: the semi-interquartile range method and the 90th percentile method (Squires et al., 2002). Meanwhile, a previous study

VALIDITY OF ASQ:SE-2-C

of the ASQ:SE, First Edition, on a Chinese sample (Bian et al., 2017) reported higher identification rates using the cutoffs based on the semi-interquartile range compared to those based on the 90th percentile. However, given the current scarcity of mental health services in China (Liu et al., 2011; Que et al., 2019), the 90th percentile method was selected for generating cutoffs. The 90th percentile scores in each ASQ:SE-2-C age interval were calculated from the national sample to serve as the cutoff points. As shown in Table 2, the proportion of children at risk (i.e., needing further evaluation and/or intervention) ranged from 9.81% to 15.86% when using the cutoff scores derived from the Chinese sample established in the current study.

[insert Table 2 here.]

Examining the Two-Factor Structure

Results from an initial analysis applying the same two-factor models as in Chen et al. (2020b) indicated poor loadings and negative item-total correlations with the remaining social competence items for the item “*Does your child seem too friendly with strangers?*” that appears in five questionnaires for 24–60 months. Therefore, in the final models, this item was revised to load on emotional rather than social competence, as suggested in an earlier study (Chen, Filgueiras, Squires, & Landeira-Fernandez, 2016). In these final models, the item “*Does your child seem too friendly with strangers?*” showed slightly higher factor loadings and no longer showed substantial negative item-total correlations; although for the 48- and 60-month intervals, this item continued to demonstrate a negligible negative item-total correlation with the total score ($r = -.01$ and $-.03$, respectively). Table 3 shows the results of these final models.

The results indicated an acceptable to good fit for the two-factor model across all ages except the two-month, 36-month, and 60-month intervals. For the six-, 12-, 18-, 24-, 30-, and 48-month age intervals, the RMSEA values were all $\leq .06$. Similarly, although some CFI and TLI

VALIDITY OF ASQ:SE-2-C

15

1
2
3 values for these intervals were $< .95$, they were all above $.90$, which may indicate their
4
5 acceptable fit with smaller sample sizes (Bentler, 1990; Weston & Gore, 2006). However, the
6
7 two-month age interval clearly demonstrated poor fit, suggesting that the two-factor model was
8
9 not appropriate at this age. On the other hand, the 36- and 60-month age intervals, although not
10
11 meeting fit criteria, approached the CFI/TLI cutoff of $> .90$ and demonstrated acceptable RMSEA
12
13 values. Further, the factor correlations tended to increase with age, reaching $\geq .80$ at many of the
14
15 older ages, indicating a lack of discrimination to some extent between these factors at older ages.
16
17 The IRT reliabilities for the social and emotional competence factors were generally acceptable
18
19 to good at $.68$ to $.82$ for all the age groups except two- ($.57$ and $.54$) and six-months ($.66$
20
21 and $.58$). The factor loadings for each age interval can be found in the supplementary materials
22
23 (Tables S2–S10).
24
25
26
27

28 [Insert Table 3 here.]
29
30

31 *Evaluating Internal Consistency*

32
33 The ordinal α ranged from $.81$ in the two- to $.94$ in the 48 month interval for the entire
34
35 ASQ:SE-2-C scale. Because of the poor fit of the two-factor model at the two-month age
36
37 interval, the ordinal α for separate social and emotional items were only calculated for the six-
38
39 month intervals and above. Among these remaining eight age intervals, the ordinal α ranged
40
41 from $.86$ in the 60- to $.94$ in the 12-month interval for the social factor and from $.77$ in the 6-
42
43 to $.90$ in the 60-month interval for the emotional factor. The specific ordinal α values for the
44
45 entire ASQ:SE-2-C and each factor (i.e., social and emotional) across age intervals are listed in
46
47 the supplementary materials (Table S1).
48
49
50

51 **Study 2**

52 **Methods**

53
54
55
56
57
58
59
60

Participants and Procedures

A regional sample of 730 children and their caregivers was recruited from 18 maternal and child healthcare clinics in Kunshan, a city of 1.65 million people in Eastern China. Recruitment flyers were distributed to caregivers of children from 15 to 72 months of age, referred for social-emotional problems from well-child check-ups, preschools, and early intervention programs. Upon consent, caregivers completed the demographic form, the ASQ:SE-2-C questionnaire, and a convergent measure questionnaire, either in a quiet room at the clinic or at home. The staff at the clinics were trained to answer questions raised by the caregivers and to provide follow-up resources, such as a diagnostic referrals or information about local early intervention programs. Table 4 summarizes the participating children's information: gender, ethnicity, disability status, mother's education, family income, and the person who completed the questionnaires. The sample size in each ASQ:SE-2-C interval ranged from 71 to 316, and the mean age of the children ranged from 17.66 (SD = 1.57) to 64.05 (SD = 4.95). Missing data ranged from 0 to 7.40% of the sample. No statistical significance was detected between those with complete data and missing data regarding the child's gender, ethnicity, disability status, mother's education, and family income. Therefore, missing data were not included in the analyses.

The assessment results for each convergent measure were compared with the ASQ:SE-2-C via the correlations of scores and the agreement of assessment categorizations (i.e., typical and at-risk). The computation of overall agreement, sensitivity, specificity, under-identification, and over-identification followed the formulas provided in *ASQ:SE-2 User's Guide* (Squires et al., 2015, p. 193). Further, descriptive, correlation, and internal consistency analyses were conducted using the Statistical Package for the Social Sciences, Version 26 (SPSS 26; IBM, 2019).

[Insert Table 4 here.]

Convergent Measures

A literature review was conducted using both English and Chinese databases to identify Chinese-language convergent measures. Three Chinese measures targeting young children's social-emotional and behavioral disorders were identified as described below.

The CITSEA was translated from the original English Infant-Toddler Social and Emotional Assessment (Carter & Briggs-Gowan, 2006) by Wang et al. (2009). Here, there are 146 items measuring four domains in social-emotional development: externalizing, internalizing, dysregulation, and competence. A three-point scale is used, where 0 refers to "not true or rarely," 1 refers to "somewhat true or sometimes," and 2 refers to "very true or often." A coding of "N" is available for some items to indicate "not applicable or no chance to observe," such as for some behaviors that occurred when switching to a new caregiver. More specifically, in the externalizing, internalizing, and dysregulation domains, higher domain scores indicate problems. However, in the competence domain, lower domain scores indicate problems, while higher scores indicate competencies. Moreover, the Wang et al. (2009) study also reported adequate to good retest reliability (.63 to .89) and split-half reliability (.55 to .90) but poor to good internal consistency (.43 to .83) as well as significant but lower-than-expected correlations with the Chinese Child Behavior Checklist for Two- and Three-year-old Children (CBCL 2–3), ranging from .23 to .49.

The original English CBCL 2–3 (Achenbach, 1992) was translated into Mandarin Chinese by Liu et al. (2003) for use in China. All 99 items measure six domains of behavioral problems: disruptive, aggressive, somatic complaints, withdrawn, depressed, and sleep problems. However, no item measures behaviors related to skills or competencies. Moreover, the Chinese

1
2
3 CBCL 2–3 uses the same three-point scale as the CITSEA, where higher scores indicate more
4
5 problems. Here, externalizing and internalizing composites can be calculated by summing the
6
7 scores of corresponding items. In Liu et al.'s (2003) study, the Chinese CBCL 2–3 showed weak
8
9 to acceptable test-retest reliability (.73 to .87), inter-rater reliability between parents and teachers
10
11 (.70 to .88), split-half reliability (.53 to .91), and internal consistency (.65 to .89).
12
13

14
15 Additionally, the Child Behavior Checklist for Four- to 18-year-old Children (CBCL 4–
16
17 18), Second Edition (Achenbach & Edelbrock, 1991), was translated and validated for use in
18
19 China as one of the earliest childhood behavior measures (Xi et al., 1992). There are 120 items
20
21 using the same three-point scale as in the CBCL 2–3, measuring nine domains: withdrawn,
22
23 somatic complaints, anxious or depressed, sexual problems, schizoid, aggressive, immature (for
24
25 boys) or obese (for girls), conduct problems (for boys) or hyperactive (for girls). The CBCL 4–
26
27 18 also allows for the computation of two composite scores: externalizing and internalizing. In
28
29 this context, a study in China (Su, Li, Wan, Yang, & Luo, 1996) reported a test-retest reliability
30
31 of .77, significantly lower scores for a known group of children with diagnosed mental disorders
32
33 than their typical peers, and significant correlations (.45 to .89) with corresponding domains on
34
35 the Chinese version of Conners' Parent Rating Scale (Goyette, Conners, & Ulrich, 1978).
36
37
38

39 *Data Analyses*

40
41
42 The correlation of scores was analyzed based on the Pearson correlation coefficients
43
44 between the ASQ:SE-2-C total scores and total or domain scores on the convergent measure.
45
46 According to Cohen (1988), in the behavioral sciences, correlation coefficients at or above .50
47
48 indicate a high correlation. Furthermore, some researchers suggest a higher threshold of .70 as
49
50 strong evidence for convergent validity (Chmielewski, Sala, Tang, & Baldwin, 2016). In the
51
52 current study, the correlation coefficients between the ASQ:SE-2-C and convergent measures
53
54
55
56
57
58
59
60

VALIDITY OF ASQ:SE-2-C

19

1
2
3 were expected to be higher than .50, preferably reaching .70. Agreement in the assessment
4
5 categorization was computed using descriptive statistics, according to the formulas provided in
6
7 *ASQ:SE-2 User's Guide* (Squires et al., 2015, p. 115). Missing data ranged from 0 to 7.40% of
8
9 the sample and were not included.
10

Results

11
12
13
14
15 Table 5 lists the correlations of the ASQ:SE-2-C total score and the two individual social
16
17 and emotional factors with the scores of the convergent measures across the six age intervals.
18
19 The results from the correlation analyses and the agreement of the classifications are described
20
21 below. These results provide evidence of the convergent properties of the ASQ:SE-2-C scores.
22
23

Correlations between the ASQ:SE-2-C and CITSEA

24
25
26 As listed in Table 5, the ASQ:SE-2-C total score showed significant correlations with all
27
28 four domains on the CITSEA except for the internalizing domain in the 30-month interval. High
29
30 correlations were found between the ASQ:SE-2-C total score and the CITSEA competence
31
32 domain, with the absolute value of the coefficients ranging from .55 to .69. The negative
33
34 correlations with the competence domain were expected because lower scores in the CITSEA
35
36 competence domain indicate more concerns. The ASQ:SE-2-C social factor showed strong
37
38 correlations with the CITSEA competence domain in all three age intervals, with the absolute
39
40 value of the coefficients ranging from .57 to .72. The ASQ:SE-2-C emotional factor showed
41
42 lower but still significant correlations with most CITSEA scores (ranging from -.33 to -.38),
43
44 except for one non-significant correlation (in the 18-month interval) with the CITSEA
45
46 competence domain.
47
48
49
50

Correlations between the ASQ:SE-2-C and CBCL 2–3

51
52 The ASQ:SE-2-C total score as well as the individual social and emotional factor scores
53
54
55
56
57
58
59
60

1
2
3 showed significant correlations with the domain and total CBCL 2–3 scores, except for the sleep
4 problems domain with the 48-month interval. The value of the significant correlation coefficients
5 ranged from .27 to .63. Relatively higher correlations seemed to be found with the CBCL 2–3
6 total score, internalizing composite, externalizing composite, and withdrawal domain. Moderate-
7 to low-level correlations (i.e., less than .50) were found with sleep problems and somatic
8 complaints.
9

10 *Correlations between the ASQ:SE-2-C and CBCL 4–18*

11 All comparisons between the ASQ:SE-2-C 60-month and CBCL 4–18 showed significant
12 correlations, with absolute coefficient values ranging from .19 to .49. Relatively lower
13 coefficients were found in somatic complaints, sexual problems, and conduct problems (for
14 boys) and hyperactivity (for girls).
15

16 [Insert Table 5 here.]
17

18 *Classification Agreements Between the ASQ:SE-2-C and the Convergent Measures*

19 As shown in Table 6, the ASQ:SE-2-C and the convergent measures showed high overall
20 agreements ranging from 76% to 85% across the six age intervals. However, the lower sensitivity
21 in the 18-month (33%) and 30-month (60%) intervals is concerning. Similarly, some concerning
22 under-identification rates were found, as high as 15% in the 18-month and 17% in the 30-month
23 intervals.
24

25 [Insert Table 6 here.]
26

27 **Discussion**

28 Healthy social-emotional development in the earliest years provides a foundation for
29 lifelong development and well-being (National Scientific Council on the Developing Child,
30 2008, 2012). Thus, the effort to measure how children gain social and emotional skills is the first
31

step in detecting those in need of additional support, intervention, or other support services. Further, the problem of unidentified mental health problems in young children requires the development of a universal screening system (Levitt, Saka, Romanelli, & Hoagwood, 2007) that in resource-limited countries, such as China and other developing countries, presents challenges. Adapting and introducing well-established measures developed in other countries has been proven to be an efficient and convenient approach to meeting the urgent needs in China. However, as Merenda (2005) pointed out, simply adopting a measure for use in another culture without systematic efforts in adaptation and validation is one of the most “ineffective and dangerous practices” (p. 322). Moreover, without proper contextualization, the results obtained by a measure can be biased and lead to incorrect interpretations.

Summary of Main Findings

We carried out two different studies in which we presented evidence of the validity of the scores of the ASQ:SE-2, a widely used parent-reported social-emotional screening measure. Specifically, the factor structure, reliability, internal consistency, and convergent validity of ASQ:SE-2-C scores were extensively examined. Several main findings were obtained. First, the adapted ASQ:SE-2-C provides culturally sensitive cutoff scores derived from a national normative sample in China. Second, the two-factor structure validated in previous studies of the ASQ:SE-2 in other countries (United States, Taiwan, and Brazil) generally showed a good fit with the Chinese national sample. Third, ASQ:SE-2 scores showed weak to acceptable evidence of convergent validity with the CITSEA, CBCL 2–3, and CBCL 4–18. These main findings suggest that the ASQ:SE-2 has the potential to be used as a screening tool to identify risks in the social-emotional development of young children in China in a timely manner.

Evidence Supporting the Two-Factor Model Fit for Older Ages

1
2
3 In Study 1, a national representative sample was established for each age interval of the
4 ASQ:SE-2-C. A two-factor structure (Chen et al., 2020b) that had been shown to be acceptable
5 with a good fit for the six-, 12-, 18-, 24-, 30-, and 48-month questionnaires was used. However,
6 fit was slightly poorer for the 36- and 60-month questionnaires with fit statistics just below the *a*
7 *priori* specified cut-offs, indicating that the two-factor structure should be explored further in
8 these age intervals in future studies. Particularly, items 3, 6, and 21 demonstrated lower loadings
9 in one or both of these age intervals (see Supplementary Materials), and this could be considered
10 in future revisions of the instrument. However, overall, the ASQ:SE-2-C showed substantial
11 consistency across age intervals in its two-factor (social and emotional) structure. These findings
12 are consistent with those from previous studies in other countries/regions (e.g., Anuniação et al.,
13 2019; Chen et al., 2016; Chen et al., 2020a; Chen et al., 2020b). This suggests that, in general,
14 the construct of social-emotional development in the early years may have some degree of
15 consistency across various cultures. However, inconsistencies exist across countries. For
16 example, the item “*Does your child seem too friendly with strangers?*” (in the 24-, 30-, 36-, 48-,
17 and 60-month questionnaires) appeared to be misfit when set to load on the social factor with the
18 U.S. sample (Chen et al., 2020 b) as well as the Chinese sample in the current study, but not with
19 a previous Taiwanese (Chen et al., 2020 a) or Brazilian sample (Anuniação et al., 2019).
20 However, a previous study (Chen et al., 2020 b) conducted differential item functioning analysis
21 and did not find significant differences between a U.S. sample and a Taiwanese sample. The
22 limited information collected on this item makes it challenging to attribute the misfit to cultural
23 differences.

24 **Relatively Weaker Model Fit in the Two-month Interval**

25 In Study 1, the sample ($N = 2,830$) was established using a stratified purposeful sampling

1
2
3 method and was representative of the Chinese population regarding geographic region and
4
5 gender (see Table 1). In the current Chinese representative sample, although the ASQ:SE-2-C
6
7 presented a generally acceptable to good model fit with the two-factor structure, the two-month
8
9 interval seems to have a weaker model fit when compared to the older age intervals. Meanwhile,
10
11 the IRT reliabilities for the social and emotional factors seemed to be lower in the younger age
12
13 intervals. Future research is needed to further examine the factor structure of the ASQ:SE-2-C
14
15 with younger children. One possible direction for future research is to investigate the extent to
16
17 which parenting experiences impact their reports on their child's social-emotional development.
18
19 This might be because many parents of two-month babies in the current study were new to the
20
21 parenting role, given the low fertility rate in China (Guo, Gietel-Basten, & Gu, 2019). Thus, they
22
23 might be struggling to differentiate between the social and emotional behaviors of their new-born
24
25 baby in the first few months of life when such behaviors are usually intertwined.
26
27
28
29

30 31 **Convergent Validity**

32
33 According to the guidelines in the Standards for Educational and Psychological Testing
34
35 (American Educational Research Association [AERA] et al., 2014), evidence for the validity of a
36
37 test can be accumulated from multiple sources. The current study examined validity evidence for
38
39 the ASQ:SE-2-C regarding the internal structure in Study 1 and relations to convergent measures
40
41 in Study 2. In Study 2, a regional sample was analyzed to examine the correlations and
42
43 agreements between the ASQ:SE-2-C and three convergent measures commonly used in
44
45 diagnosing social-emotional and behavioral problems in China. The ASQ:SE-2-C total score as
46
47 well as the social and emotional sub-scale scores presented statistically significant correlations
48
49 with most domain/total scores and the three convergent measures, although the magnitude of
50
51 correlation ranged from weak (e.g., $r = .18$ between the 60-month ASQ:SE-2-C and somatic
52
53
54
55
56
57
58
59

VALIDITY OF ASQ:SE-2-C

24

1
2
3 complaints subdomain in the CBCL 4-18) to strong (e.g., $r = .72$ between the 30-month ASQ:SE-
4 2-C and competence domain in the CITSEA).

5
6
7
8 In this context, the varied values of correlation coefficients may be explained by the
9 structure of the ASQ:SE-2-C. In addition to the scored items, the ASQ:SE-2-C includes several
10 overall open-ended questions at the end of each questionnaire. These overall questions are not
11 scored, and thus they were not included in the current study; however, they collect information to
12 supplement the scored items. For example, in the 36-month questionnaire, an overall question
13 asks, “*Do you have concerns about your child’s eating, sleeping, or toileting habits? If yes,*
14 *please explain.*” Here, if not in the ASQ:SE-2-C scored items, sleep problems may be picked up
15 in parents’ responses to this overall question. Moreover, as a parent-completed measure
16 developed for screening purposes, it is parent friendly and cost efficient. The ASQ:SE-2-C
17 features only 31 (18-month) to 36 (60-month) scored items, which is much shorter than the 146
18 items on the CITSEA, 99 items on the CBCL 2–3, and 120 items on the CBCL 4–18. The overall
19 ASQ:SE-2-C questions are also crucial in identifying concerns not captured in the scored items.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 Classification agreement between the ASQ:SE-2-C and the four converging measures
36 also varied across the six age intervals. The low sensitivity (33% in the 18-month; 60% in the 30-
37 month) and high under-identified rate (15% in 18-month; 17% in 30-month) results are
38 particularly concerning. Here, it may be necessary to triangulate results from the scoring items
39 on the ASQ:SE-2-C with information collected from other sources to strengthen the decision on
40 whether a child needs to be referred for further evaluation. However, according to the
41 instructions provided in *ASQ:SE-2 User’s Guide* (Squires et al., 2015), a referral for further
42 evaluation and/or intervention services should not be made solely dependent on the ASQ:SE-2
43 total score. Rather, parental concerns indicated in the comment areas in the scored item section
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

VALIDITY OF ASQ:SE-2-C

25

1
2
3 and in the overall questions section should be addressed, and a referral should be made based on
4
5 these concerns, even if the ASQ:SE-2-C total score is below the cutoff. Future research is needed
6
7 to investigate whether parents' responses to the overall questions actually enhance the sensitivity
8
9 of the ASQ:SE-2-C.
10

11
12 One of the biggest challenges in considering evidence for the validity of ASQ:SE-2-C
13
14 scores for use in China came from the lack of convergent measures. Compared to the CITSEA,
15
16 CBCL 2–3, and CBCL 4–18, the stronger emphasis on strength-based behaviors in the ASQ:SE-
17
18 2-C reflects a paradigm shift away from the traditional emphasis on deficits and problems to
19
20 target a more competence-based sample of behaviors (Abrahams et al., 2019). This strengths-
21
22 based or competence-based feature of the ASQ:SE-2-C could potentially make it a more socially
23
24 valid tool since it is considered more friendly and acceptable among early childhood educators
25
26 and parents (Pontoppidan et al., 2017). Further, compared to the more problem-based measures,
27
28 the ASQ:SE-2-C may be a more socially valid option for universal screening efforts in China,
29
30 where limited professional resources make it critical to involve educators and parents in
31
32 screening. One possible direction for future research is to compare the ASQ:SE-2-C with other
33
34 measures that include strengths-based items, such as the Strengths and Difficulties Questionnaire
35
36 (Goodman, 2001).
37
38
39
40
41

42 In addition to the different emphases and sampled behaviors between the ASQ:SE-2-C
43
44 and convergent measures, the findings in Study 2 also drew attention to the overall questions
45
46 section, for which analyses are intended to be performed in future studies. As a parent-completed
47
48 screening measure, the relatively short length of the ASQ:SE-2-C was designed so that the
49
50 overall open-ended questions section could capture a wide range of parental concerns about
51
52 children's social-emotional development. Further research is needed to describe how Chinese
53
54
55
56
57
58
59
60

1
2
3 practitioners interpret and address parental concerns in the ASQ:SE-2-C overall questions
4
5 section and the validity of the ASQ:SE-2-C referral procedure that requires the consideration of
6
7 both the ASQ:SE-2-C total score and the parental concerns recorded in the overall questions
8
9 section.
10

11
12 Although the current study supports the reliability and validity of ASQ:SE-2-C scores for
13
14 the screening of children using an instrument with parent- and early childhood educator-friendly
15
16 features, future research is needed to inform the selection of measures and the interpretation of
17
18 the screening results. For example, further investigations are needed to understand how the
19
20 ASQ:SE-2 items function differently across different countries and how the two-factor structure
21
22 of the ASQ:SE-2 performs across the different ages.
23
24

25 26 **Limitations** 27

28
29 The limitations of the current study can be attributed to the participant samples and
30
31 convergent measures. In the development of the national Chinese representative sample, a
32
33 stepwise procedure was used to mimic the distribution in the national census data in terms of
34
35 geographic regions and rural and urban types of family registration. However, the national
36
37 normative sample was found to over-represent children with rural family registrations. This is
38
39 partly due to the rapid growth in urban populations at the time of the normative sampling. The
40
41 sample plan was made based on a much higher rural population reported in 2014. In addition, the
42
43 national sample also over-represented children whose mothers had higher educational attainment
44
45 (high school or college and above) and children of a minority ethnicity. In this context, caution is
46
47 needed when using the ASQ:SE-2-C cutoffs with children whose mothers have lower-than-high-
48
49 school education and those from non-Han ethnic groups. Future research should examine how
50
51 mothers' educational attainment impacts children's scores on the ASQ:SE-2-C to better inform
52
53
54
55
56
57
58
59
60

1
2
3 the users of this questionnaire.
4

5 **Conclusion**

6
7 Findings from the current study suggest that the ASQ:SE-2-C is a psychometrically
8 sound measure for early childhood social-emotional screening in China. The parent-completed
9 design and strengths-based items indicate the potential of the ASQ:SE-2-C to be a cost-effective
10 tool for universal screening and timely identification of young children who need evaluation and
11 intervention services. Parents should be actively involved in the screening of their children by
12 responding to the scored items and the open-ended questions in the overall questions section in
13 the ASQ:SE-2-C.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- Abrahams, L., Pancorbo, G., Primi, R., Santos, D., Kyllonen, P., John, O. P., & De Fruyt, F. (2019). Social-emotional skill assessment in children and adolescents: Advances and challenges in personality, clinical, and educational contexts. *Psychological Assessment, 31*(4), 460-473. <https://doi.org/10.1037/pas0000591>
- Achenbach, T. M. (1992). *Manual for child behavior checklist/2-3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., & Edelbrock, C. (1991). *Child behavior checklist*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.
- American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anuniação, L., Squires, J., Clifford, J., & Landeira-Fernandez, J. (2019). Confirmatory analysis and normative tables for the Brazilian Ages and Stages Questionnaires: Social-emotional. *Child: Care, Health and Development, 45*(3), 387-393. <https://doi.org/10.1111/cch.12649>
- Bagner, D. M., Rodríguez, G. M., Blake, C. A., Linares, D., & Carter, A. S. (2012). Assessment of behavioral and emotional problems in infancy: A systematic review. *Clinical Child and Family Psychology Review, 15*(2), 113-128. <https://doi.org/10.1007/s10567-012-0110-2>
- Bayley, N. (2006). *Bayley Scales of Infant Development and Toddler Development: Technical manual*. San Antonio, TX: The Psych Corp.

- 1
2
3 Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*,
4
5 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
6
7
- 8 Bian, X., Xie, H., Squires, J., & Chen, C.Y. (2017). Adapting a parent-completed, social-
9
10 emotional questionnaire in China: The Ages & Stages Questionnaires: Social-Emotional.
11
12 *Infant Mental Health Journal*, 38(2), 258–266. <https://doi.org/10.1002/imhj.21626>
13
14
- 15 Bricker, D., Macy, M., Squires, J., & Marks, K. (2013). *Developmental screening in your*
16
17 *community: An integrated approach for connecting children with services*. Baltimore, MD:
18
19 Paul H. Brookes Publishing.
20
21
22
- 23 Brown, T. A. (2006). *Confirmatory factor analysis for applied researchers*. New York, NY: The
24
25 Guilford Press.
26
27
- 28 Cai, L., & Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item
29
30 factor models. *British Journal of Mathematical and Statistical Psychology*, 66(2), 245–
31
32 276. <https://doi.org/10.1111/j.2044-8317.2012.02050.x>
33
34
- 35 Campos, J. J., Mumme, D. L., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective
36
37 on the nature of emotion. *Monographs of the Society for Research in Child Development*,
38
39 59(2-3), 284–303. <https://doi.org/10.2307/1166150>
40
41
42
- 43 Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2019). Malleability, plasticity, and
44
45 individuality: How children learn and develop in context. *Applied Developmental Science*,
46
47 23(4), 307–337. <https://doi.org/10.1080/10888691.2017.1398649>
48
49
- 50 Carter, A. S., Briggs-Gowan, M. J., & Davis, N. O. (2004). Assessment of young children's social-
51
52 emotional development and psychopathology: Recent advances and recommendations for
53
54 practice. *Journal of Child Psychology and Psychiatry*, 45(1), 109–
55
56
57
58
59

1
2
3 134. <https://doi.org/10.1046/j.0021-9630.2003.00316.x>
4
5

6 Carter, A. S., & Briggs-Gowan, M. J. (2006). *The Infant-Toddler Social & Emotional Assessment*
7
8 *(ITSEA)*. San Antonio, TX: Psychological Corporation, Harcourt Assessment.
9

10
11 Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R
12
13 environment [R package]. *Journal of Statistical Software*, 48(6), 1–29.
14

15
16 <https://doi.org/10.18637/jss.v048.i06>
17

18
19 Chen, C. Y., Filgueiras, A., Squires, J., & Landeira-Fernandez, J. (2016). Examining the factor
20
21 structure of an early childhood social emotional screening assessment. *The Journal of*
22
23 *Special Education and Rehabilitation*, 17(3-4), 89–104.
24

25
26 <https://doi.org/10.19057/jser.2016.12>
27

28
29 Chen, L., Li, H., Wei, W., & Zhang, Y. (2015). 广西柳州市城区 2 208 例 2~3 岁儿童行为问题
30
31 调查分析[Study on the prevalence of behavioral problems among 2,208 children aged 2-3
32
33 years and from Liuzhou City]. *Chinese Journal of Child Health Care*, 23(3), 321–323.
34

35
36
37 Chen, C. Y., Squires, J., Chen, C. I., Wu, R., & Xie, H. (2020a). The adaptation and psychometric
38
39 examination of a social-emotional developmental screening tool in Taiwan. *Early*
40
41 *Education and Development*, 31(1), 27–46.
42

43
44 <https://doi.org/10.1080/10409289.2019.1611126>
45

46
47 Chen, C. Y., Squires, J., & Scalise, K. (2020b). Evaluating the dimensionality and psychometric
48
49 properties of a social-emotional screening instrument for young children. *Infants & Young*
50
51 *Children*, 33(2), 142–159. <https://doi.org/10.1097/IYC.000000000000163>
52

53
54 China National Bureau of Statistics. (2017). China statistical yearbook 2017. Retrieved from

55
56 <http://www.stats.gov.cn/tjsj/ndsj/2017/indexeh.htm>
57
58
59

- 1
2
3 China National Health Commission. (2019). Announcement of healthy China movement – Action
4 plan for children and youth mental health actions (2019-2022). Retrieved from
5
6 http://www.gov.cn/xinwen/2019-12/27/content_5464437.htm
7
8
9
10 Chmielewski, M., Sala, M., Tang, R., & Baldwin, A. (2016). Examining the construct validity of
11 affective judgments of physical activity measures. *Psychological Assessment, 28*(9), 1128–
12 1141. <https://doi.org/10.1037/pas0000322>
13
14
15
16
17
18 Chung, K. M., Jung, W., Yang, J. W., Ben-Itzhak, E., Zachor, D. A., Furniss, F., ... & Barker, A.
19 A. (2012). Cross cultural differences in challenging behaviors of children with autism
20 spectrum disorders: An international examination between Israel, South Korea, the United
21 Kingdom, and the United States of America. *Research in Autism Spectrum Disorders, 6*(2),
22 881–889. <https://doi.org/10.1016/j.rasd.2011.03.016>
23
24
25
26
27
28
29
30 Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ:
31 Lawrence Earlbaum Associates.
32
33
34
35 Denham, S. A., Wyatt, T., Bassett, H. H., Echeverria, D., & Knox, S. S. (2009). Assessing social-
36 emotional development in children from a longitudinal perspective. *Journal of*
37 *Epidemiology and Community Health, 63*(1), 37–52.
38
39 <https://doi.org/10.1136/jech.2007.070797>
40
41
42
43
44
45 Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus,*
46 *2*(222), 1–17. <https://doi.org/10.1186/2193-1801-2-222>
47
48
49
50 Drotar, D., Stancin, T., & Dworkin, P. (2008). *Pediatric developmental screening: Understanding*
51 *and selecting screening instruments*. National Academy for State Health Policy. Retrieved
52 from <https://healthychild.nashp.org/pediatric-developmental-screening-understanding-and->
53
54
55
56
57
58
59
60

[selecting-screening-instruments/](#)

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods, 14*(3), 275–299.

<https://doi.org/10.1037/a0015825>

Gadermann, A. Guhn, M., & Zumbo, B. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation, 17*(3), 1–3. <https://doi.org/10.7275/n560-j767>

Glascoe, F., & Robertshaw, N. S. (2007). New AAP policy on detecting and addressing developmental and behavioral problems. *Journal of Pediatric Health Care, 21* (6), 407–412. <https://doi.org/10.1016/j.pedhc.2007.08.008>

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1337–1345. <https://doi.org/10.1097/00004583-200111000-00015>

Goyette, C. H., Conners, C. K., & Ulrich, C. F. (1978). Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology, 6*, 221–236. <https://doi.org/10.1007/BF00919127>

Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer. <https://doi.org/10.1007/978-1-4614-4018-5>

Guo, Z., Gietel-Basten, S., & Gu, B. (2019). The lowest fertility rates in the world? Evidence from the 2015 Chinese 1% sample census. *China Population and Development Studies, 2*(3), 245–258.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:

Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

IBM. (2019). IBM SPSS statistics for Windows (version 26) [Computer software]. Armonk, NY: IBM Corp. Retrieved from <https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-26>

International Test Commission. (2005). ITC guidelines for translating and adapting tests. Retrieved from https://www.intestcom.org/files/guideline_test_adaptation.pdf

International Test Commission. (2017). The ITC guidelines for translating and adapting tests (2nd ed.). Retrieved from <http://www.intestcom.org/>

Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies and commentaries. *Social Policy Report*, 26(4), 1–33. <https://doi.org/10.1002/j.2379-3988.2012.tb00073.x>

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45(2), 163–191. <https://doi.org/10.1016/j.jsp.2006.11.005>

Liang, J., Wang, Z., Li, Y., Li, H., Qi, J., H., X., & Bai, R. (2020). 新型冠状病毒肺炎流行期间学龄前儿童心理健康状况调查及影响因素分析 [Investigation on the mental health of preschool children and the relevant factors during corona virus disease 2019]. *Chinese Journal of Child Health Care*, 17–21. Advance online publication.

Liu, L., Wu, L., & Yao, K. (2003). 2 ~ 3 岁幼儿行为量表全国城市常模的制定 [Institution of child behavior checklist (CBCL) norm for 2 to 3 years children in national cities]. *Chinese Journal of Child Health Care*, 11(6), 377–379.

VALIDITY OF ASQ:SE-2-C

34

1
2
3 Liu, J., Ma, H., He, Y. L., Xie, B., Xu, Y. F., Tang, H. Y., ... Yu, X. (2011). Mental health system
4 in China: History, recent service reform and future challenges. *World Psychiatry, 10*(3),
5 210–216. <https://doi.org/10.1002/j.2051-5545.2011.tb00059.x>
6
7

8
9
10 Matson, J. L., Worley, J. A., Fodstad, J. C., Chung, K. M., Suh, D., Jhin, H. K., ... & Furniss, F.
11 (2011). A multinational study examining the cross cultural differences in reported
12 symptoms of autism spectrum disorders: Israel, South Korea, the United Kingdom, and the
13 United States of America. *Research in Autism Spectrum Disorders, 5*(4), 1598–1604.
14
15
16
17
18
19
20 <https://doi.org/10.1016/j.rasd.2011.03.007>
21

22
23 Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in
24 multidimensional contingency tables. *Psychometrika, 71*, 713.
25
26
27 <https://doi.org/10.1007/s11336-005-1295-9>
28

29
30 Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis.
31
32 *Multivariate Behavioral Research, 49*(4), 305–328.
33
34
35 <https://doi.org/10.1080/00273171.2014.911075>
36

37 Merenda, P.F. (2005). Cross-cultural adaptation of educational and psychological testing. In R. K.
38 Hanbleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and*
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
psychological tests for cross-cultural assessment (pp. 321-342). New York, NY:
Psychology Press Taylor & Francis Group.

National Scientific Council on the Developing Child. (2008/2012). *Establishing a level foundation*
for life: Mental health begins in early childhood: Working paper No. 6. Updated edition.
Retrieved from www.developingchild.harvard.edu

Ning, M., He, H., & Yu, M. (2017). 儿童心理行为问题及影响因素概述 [Review of children

mental health problem and influencing factors]. *Chinese Journal of Women and Children Health*, 8(3), 4–6.

- Pang, Y., Lee, C. M., Wright, M., Shen, J., Shen, B., & Bo, J. (2018). Challenges of case identification and diagnosis of autism spectrum disorders in China: A critical review of procedures, assessment, and diagnostic criteria. *Research in Autism Spectrum Disorders*, 53, 53–66. <https://doi.org/10.1016/j.rasd.2018.06.003>
- Pizur-Barnekow, K., Erickson, S., Johnston, M., Bass, T., Lucinski, L., & Bleuel, D. (2010). Early identification of developmental delays through surveillance, screening, and diagnostic evaluation. *Infants & Young Children*, 23(4), 323–330. <https://doi.org/10.1097/IYC.0b013e3181f422a4>
- Pontoppidan, M., Niss, N. K., Pejtersen, J. H., Julian, M. M., & Vaever, M. S. (2017). Parent report measures of infant and toddler social-emotional development: A systematic review. *Family Practice*, 34(2), 127–137. <https://doi.org/10.1093/fampra/cmz003>
- Powell, G., Mackrain, M., & LeBuffe, P. (2007). *Devereux early childhood assessment for infants and toddlers technical manual*. Lewisville, NC: Kaplan Early Learning Corporation.
- Que, J., Lu, L., & Shi, L. (2019). Development and challenges of mental health in China. *General Psychiatry*, 32(1), e100053. <https://doi.org/10.1136/gpsych-2019-100053>
- Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized conditional SEM: A case for conditional reliability. *Applied Psychological Measurement*, 31(3), 169–180. <https://doi.org/10.1177/0146621606291569>
- Raver, C. C., & Zigler, E. F. (1997). Social competence: An untapped dimension in evaluating head start's success. *Early Childhood Research Quarterly*, 12(4), 363–385.

[https://doi.org/10.1016/S0885-2006\(97\)90017-X](https://doi.org/10.1016/S0885-2006(97)90017-X)

R Core Team. (2019). R: A language and environment for statistical computing (version 3.6.1)

[Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved

February 10, 2020, from <http://www.r-project.org/index.html>

Reid, K. B., Sacrey, L. A. R., Zwaigenbaum, L., Raza, S., Brian, J., Smith, I. M., ... &

Vaillancourt, T. (2020). The association between social emotional development and

symptom presentation in autism spectrum disorder. *Development and Psychopathology*,

32(4), 1206–1216. <https://doi.org/10.1017/S0954579420000711>

Revelle, W. (2019). psych: Procedures for psychological, psychometric, and personality research

(version 1.9.12) [R package]. Evanston, IL: Northwestern University. Retrieved February

10, 2020, from <https://CRAN.R-project.org/package=psych>

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores.

Psychometrika Monograph Supplement, 34(4), 100.

Sandall, S., Hemmeter, M. L., Smith, B. J., & McLean, M. E. (2005). *DEC recommended*

practices: A comprehensive guide for practical application in early intervention/early

childhood special education. Longmont, CO: Sopris West.

Squires, J., Bricker, D., & Twombly, E. (2002). *Ages & Stages Questionnaires, Social-Emotional*

(ASQ:SE): A parent-completed, child-monitoring system for social-emotional behaviors.

Baltimore, MD: Paul H. Brookes Publishing.

Squires, J., Bricker, D., & Twombly, E. (2015). *Ages & Stages Questionnaires: Social-Emotional*

(2nd ed.). Baltimore, MD: Paul H. Brookes Publishing.

Su, L., Li, X., Wan, G., Yang, Z., & Luo, X. (1996). [The norms of Achenbach Child Behavior

Checklist in Hunan province]. *Chinese Journal of Clinical Psychology*, 4(1), 24–28.

The World Bank. (2021). Population, total – China. Retrieved from

<https://data.worldbank.org/indicator/SP.POP.TOTL?locations=CN>

Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review.

Assessment, 18(3), 291–307. <https://doi.org/10.1177/1073191110374797>

United Nations International Children’s Emergency Fund (UNICEF). (2016). Population status of children in China in 2015: Facts and figures. Retrieved from

<http://archive.unicef.cn/cn/uploadfile/2017/1009/20171009112641471.pdf>

Wang, H., Zhang, J., Huang, X., Liu, G., Lian, G., & Shi, S. (2009). 中国城市幼儿情绪及社会性发展量表标准化的信度和效度分析 [Reliability and validity of the standardized Chinese version of the urban infant-toddler social and emotional assessment (CITSEA)]. *Chinese Journal of Child Health Care*, 17(3), 271–274.

Weston, R., & Gore, P. A., Jr. (2006). A brief guide to structural equation modeling. *The*

Counseling Psychologist, 34(5), 719–751. <https://doi.org/10.1177/0011000006286345>

Xi, R., Tang, H., Zhang, Z., Cai, X., Cheng, Z., Yu, G., & Jiang, Z. (1992). 全国 22 个省市 26 个

单位 24013 名城市在校少年儿童行为问题调查——独生子女精神卫生问题的调查、

防治和 Achenbach’s 儿童行为量表中国标准化[A survey of the problem behaviors on

24,013 urban students from 26 units in 22 provinces and cities in China]. *Shanghai*

Archives of Psychiatry, 4(1), 47–55.

Xie, H., Bian, X., Chen, C. Y., Squires, J., & Lu, P. (2019). Examining the convergent evidence of

a parent-completed, social-emotional screening tool in China. *Journal of Child and Family*

1
2
3 *Studies*, 28(6), 1471–1480. <https://doi.org/10.1007/s10826-019-01362-3>

4
5
6 Yates, T., Ostrosky, M. M., Cheatham, G. A., Fettig, A., Shaffer, L., & Santos, R. M. (2008).

7
8 Research synthesis on screening & assessing social-emotional competence. *Center on the*
9
10 *Social Emotional Foundations for Early Learning*. Retrieved from

11
12 http://csefel.vanderbilt.edu/documents/rs_screening_assessment.pdf

13
14
15
16 Ye, S., & Tan, D. (2015). 学龄前儿童行为问题的现状及影响因素分析 [A survey of preschool
17
18 child behavior problems]. *Modern Primary and Secondary Education*, 31(4), 93–100.

19
20
21 Zhu, J., & Zhang, J. (2008). Contemporary trends and developments in early childhood education
22
23 in China. *Early Years*, 28(2), 173–182. <https://doi.org/10.1080/09575140802163584>