Autism spectrum disorder screening in Chinese-language preschools

Kathy Kar-man Shum[1] iD **, Rose Mui-fong Wong[2],**
Angel Hoe-chi Au[3] and Terry Kit-fong Au[1]

Abstract

This study examined the reliability and validity of the 13-item Classroom Observation Scale as used by teachers and non-clinically trained observers to identify children who more likely than their peers to have autism spectrum disorder in less-resourced preschools. A total of 534 children (ages 2;10 to 4;5, Mean = 3;8) from nine Chinese-language preschools serving families from lower-middle to middle socioeconomic backgrounds in Hong Kong were observed in their first preschool year using the Classroom Observation Scale. The 75 screen-positive children and 55 randomly selected typically developing peers were clinically assessed for autism spectrum disorder 1 year later. The Classroom Observation Scale as used by teachers and non-clinically trained researchers helped to identify preschoolers who were later diagnosed with autism spectrum disorder with odds ratios of 3.11 and 8.66, respectively. This study provided further evidence on the versatility and ecological validity of the Classroom Observation Scale for use by preschool teachers and observers with little or no clinical training in the early identification of children with autism spectrum disorder in community settings.

Lay abstract

The 13-item Classroom Observation Scale is an autism spectrum disorder screening tool for teachers and non-clinically trained observers to make real-time observation of children's peer interaction (or the lack thereof) in regular preschool classrooms. The Classroom Observation Scale was originally developed in English and validated with ethnically diverse preschoolers at English-speaking international schools serving families from middle to middle-upper socioeconomic backgrounds in Hong Kong. These private schools can usually afford a higher teacher–student ratio, which is not typical for most preschools. This study, therefore, investigated whether the Classroom Observation Scale is ecologically valid when used by Chinese teachers with teacher–student ratios typically found in less-resourced preschools. We found that the Classroom Observation Scale reliably helped observers with little or no clinical training—research assistants with just a few hours of Classroom Observation Scale training and preschool teachers with an hour of briefing—to identify children in their first year of Chinese-language preschool who were more likely than their peers to have autism spectrum disorder. Reliability estimates of Classroom Observation Scale-Teacher and Classroom Observation Scale-Researcher in this study were comparable to those for the original English Classroom Observation Scale. Our results provided further evidence on the versatility and ecological validity of the Classroom Observation Scale for use by preschool teachers and non-clinically trained observers in the early identification of children with autism spectrum disorder in community settings.

Keywords

autism spectrum disorder, Chinese, Classroom Observation Scale, preschoolers, screening

Early intervention for children with autism spectrum disorder (ASD) requires early identification. Au et al. (2021) developed the 13-item Classroom Observation Scale (COS) as a valid and reliable ASD screening tool for teachers and non-clinically trained observers to make real-time observation of children's peer interaction (or the lack thereof) in regular preschool classrooms. It focuses on

[1]The University of Hong Kong, China
[2]WAY Psychological Services, China
[3]Autism Partnership Hong Kong, China

Corresponding author:
Kathy Kar-man Shum, Department of Psychology, The University of Hong Kong, Pokfulam Road, Hong Kong, China.
Email: kkmshum@hku.hk

peer interaction because children with ASD often display more noticeable problems (compared to non-ASD children) when they are around peers without adults hovering around to scaffold and instruct (Corbett, Schupp, Simon, Ryan, & Mendoza, 2010). The COS hence contrasts substantially with well-known ASD screening tools for preschoolers.

For example, both the 12-item *Diagnostic and Statistical Manual of Mental Disorders* (DSM)-Autism Spectrum Problems Scale (*DSM*-ASD Scale; Achenbach, 2014) from the Child Behavior Checklist for Ages 1½–5 years (CBCL/1½–5) and its Caregiver-Teacher Report Form (C-TRF; Achenbach & Rescorla, 2000) consist of seven items on social communication/interaction (SCI) and five items on restricted interests, repetitive behaviors (RRB; Rescorla, Ghassabian, et al., 2019). By contrast, the 13-item COS consists of two items on RRB and one on self-regulation challenge, while the rest of the items are on social interaction. Moreover, the SCI items of the C-TRF *DSM*-ASD Scale mostly describe social responding behaviors (i.e. whether the child responds to others' initiation of interaction), whereas the COS items focus more on social initiation behaviors (e.g. "Initiates to point out things in the environment to other children or adults," "Initiates conversation with other children"). Note that social initiation behaviors, relative to social responding (or non-responding) behaviors, might be more readily noticed by teachers and others in a busy preschool classroom. The COS proves to have good psychometric properties and, crucially, strong predictive validity in identifying preschoolers under the age of 4½ years to be more likely than their peers to have ASD diagnosable about 1½ years later (Au et al., 2021).

The COS was developed in English and validated with ethnically diverse preschoolers at English-speaking international schools serving families from middle to middle-upper socioeconomic backgrounds in Hong Kong. These private schools can usually afford a teacher–student ratio ranging from about 1:8 to 1:12. Such manpower ratio is not typical for most preschools in Hong Kong, or other cities in China or Asia (Education Bureau, Hong Kong SAR, 2020; Li, Rao, & Tse, 2012; Peach, 1994). Moreover, most teachers at English-speaking international preschools in Hong Kong are expatriates from North America, Europe, and Australia. By contrast, the majority of teachers in Chinese-language preschools are Hong Kong Chinese trained locally. Such variations in teacher–student ratio and potential cultural differences in teachers' sensitivity about ASD symptomatology could affect teachers' use of observation scales (Rescorla, Given, Glynn, Ivanova, & Achenbach, 2019). This study, therefore, investigated whether the COS is ecologically valid when used by Chinese teachers with teacher–student ratios typically found in less-resourced preschools.

Method

Participants

This study was approved by the Human Research Ethics Committee of the authors' university. Parents gave written consent for 534 children (age 2;10 to 4;5, Mean = 3;8, SD = 4 months, with 273 boys and 261 girls): (1) to be observed in nine Chinese-language preschools serving families from lower-middle to middle socioeconomic backgrounds in Hong Kong, with eight of these preschools qualified for direct subsidy from the government, and (2) to participate, if selected, in ASD clinical assessment 1 year later. The children's parents and preschool teachers ($n = 40$) also gave written consent to provide information about the children.

Procedure

The procedure was generally similar to that of Au et al. (2021). Parents were invited to participate about 5 months after their children had started preschool. An experienced clinical psychologist trained four research assistants (with university-level psychology coursework but no prior clinical training) to use the 13-item COS. Good interrater reliability was achieved after about 6 h of training. The raters then observed each child participant on two school days no more than 35 days apart (Mean = 7.3 days; SD = 5.9 days), with four to seven children per school day in random order for each round of 1 min observations. Each target child was observed for about 30 to 40 one-minute intervals in total. Altogether, 40 teachers from the nine preschools rated their students using COS after attending a 30- to 45-min briefing on the scoring of the checklist items.

In the second semester of the children's first preschool year, we identified children more likely than their peers to have ASD using two approaches (Figure 1): (1) bottom 15% on COS-Teacher and below the median (*Mdn* = 40) on COS-Researcher ($n = 64$), (2) bottom 15% on COS-Researcher and below the median (*Mdn* = 34) on COS-Teacher ($n = 56$). We identified 75 out of 534 children as more likely to have ASD, noting considerable overlap of screen-positives between the two approaches. They were mixed in with 55 randomly selected screen-negatives (typically developing control) for ASD assessment about 1 year after the COS data collection (i.e. in the second semester of second year in preschool). The clinical assessments of these 130 children—75 screen-positives and 55 controls—were done by a clinical psychologist with about 10 years of clinical experience in public hospitals and private practice, with formal Autism Diagnostic Observation Schedule–2 (ADOS-2) and Autism Diagnostic Interview–Revised (ADI-R) training for clinical purposes and kept blind to the children's screen-positive versus control status. ADOS-2 was administered to the children and their parents were interviewed about the children's developmental history using ADI-R.
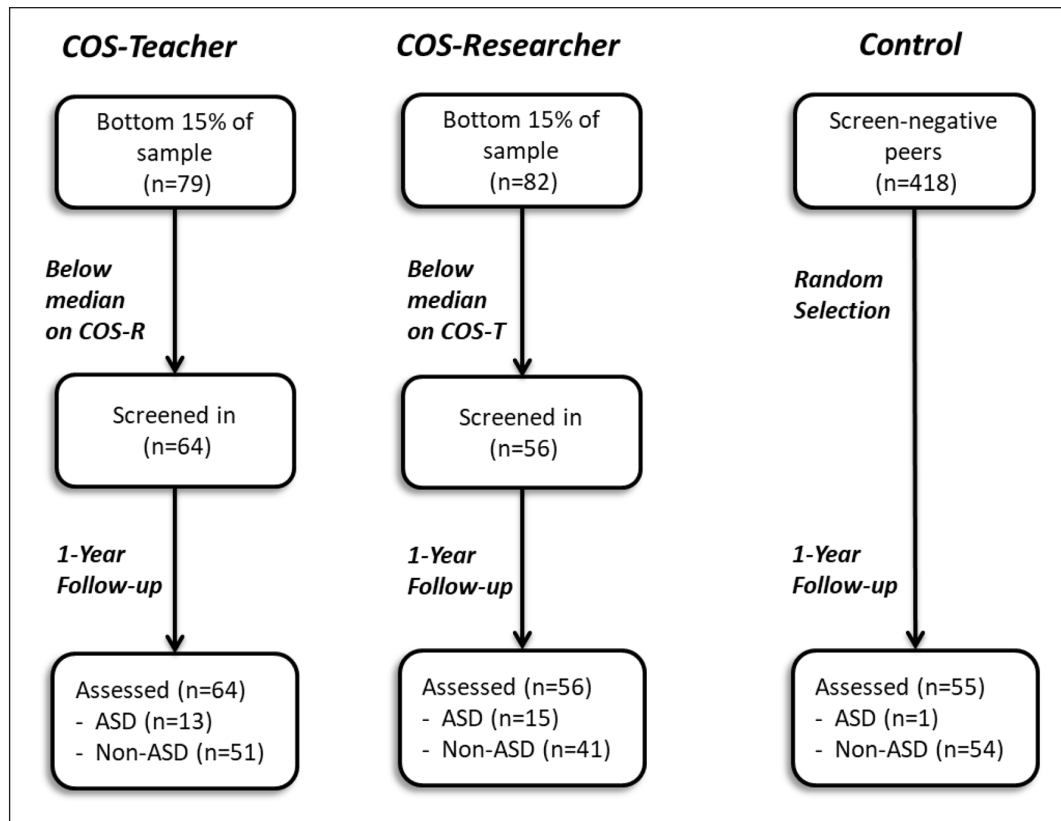
Figure 1. Two screening approaches for ASD: 75 out of 534 children were identified as more likely to have ASD, with considerable overlap of screen-positives between the two approaches. The screen-negative children ($n = 418$) were above the bottom 15% cutoff on both COS-Teacher and COS-Researcher. Teachers and researchers rated children using the COS in the first year of preschool, and ADOS-2 assessments were conducted at the 1-year follow-up.

Instruments

COS. The 13-item COS used in Au et al. (2021) was translated and back-translated by our Chinese–English bilingual research team. Translation fidelity and quality were checked, and the scale was copyedited by experts in child clinical psychology and developmental psychology. (COS-Chinese is available upon request for research purposes.) It contains 10 items on challenges in peer interaction (e.g. "Directs facial expressions to peers"), 2 on RRB (e.g. "Engages in repetitive behaviors or unusual mannerisms"), and 1 on self-regulation challenge (e.g. "Sits down or stays seated during structured teaching times"). Each item was rated on a 5-point scale (1 = very rarely or never; 2/3/4/5 = less often than/ about as often as/ more often than/ much more often than most students, respectively). The maximum total score is 65, with lower COS scores indicating more problem behaviors observed.

ADOS-2. This is a semi-structured, standardized tool for autistic disorder and ASD (Lord et al., 2012; Oosterling et al., 2010). It provides opportunities for children to engage in communication, social interaction, and play (or imaginative use of materials). All 130 children, who spoke in multi-word utterances, were assessed with Module 2 of ADOS-2.

ADI-R. This interview protocol for parents of children aged 2 years or above (Kim & Lord, 2012) was used to supplement ADOS-2.

Community Involvement: Community service providers for ASD are involved in this study.

Results

Reliabilities

Cronbach's alpha was 0.90 for both COS-Teacher and COS-Researcher, indicating good internal reliability. Intraclass correlation coefficients (ICCs; mean-rating ($k = 2$), absolute-agreement, two-way random-effects model) for inter-rater and test–retest reliabilities for COS-Researcher were 0.85 and 0.73, respectively (Koo & Li, 2016) and comparable to those for the English COS (Au et al., 2021).

Validity of the screening

Among the 130 preschoolers (ages 4;1 to 5;9, Mean = 4;8, SD = 4 months; 75 boys and 55 girls) assessed for ASD on average about 1 year after the COS data collection (10 to 19 months, Mean = 13.4 months, SD = 1.6 months), 18
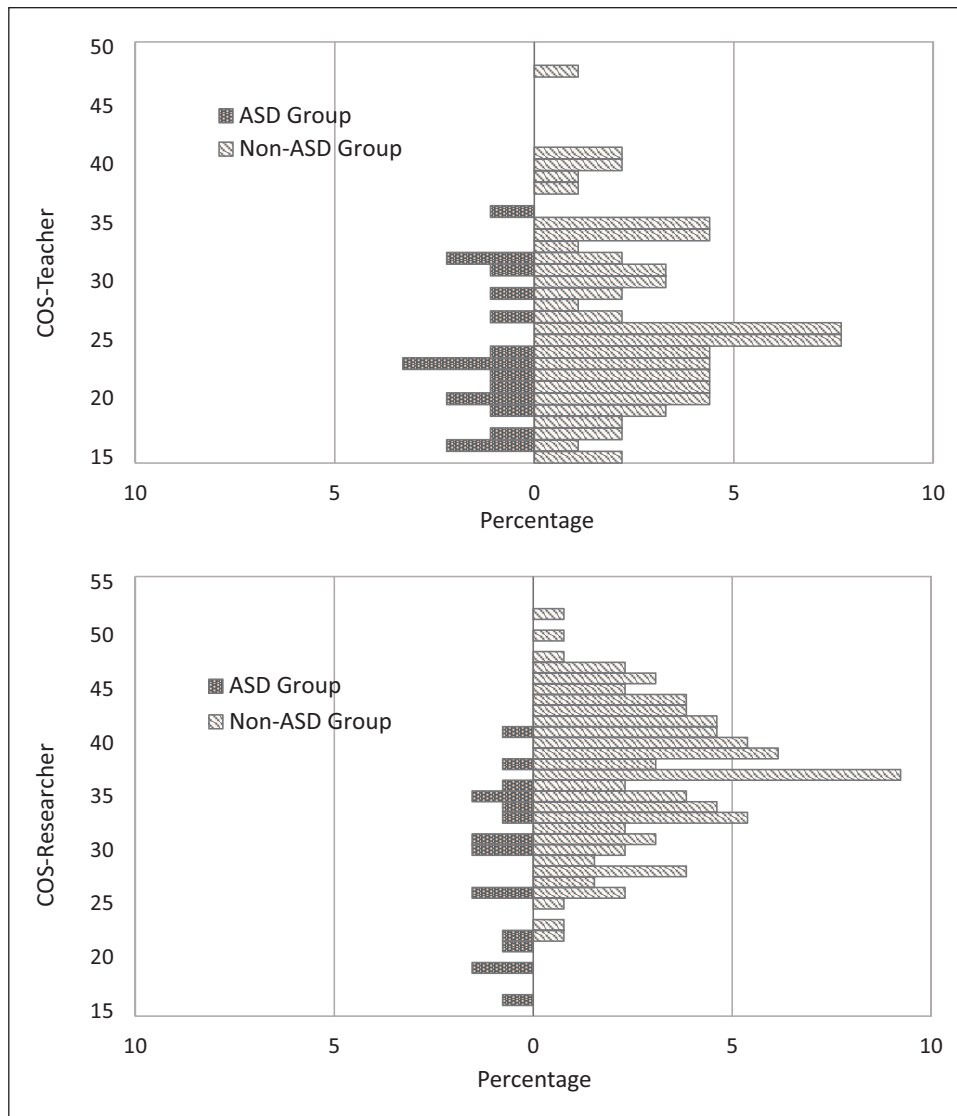
Figure 2. Relative frequency distribution of scores on COS-Teacher (top) and COS-Researcher (bottom) for the ASD and non-ASD groups.

(16 boys and 2 girls; ages 4;2 to 5;5, Mean = 4;8, SD = 4.4 months) met the diagnostic criteria (ASD group), whereas the remaining 112 children did not (non-ASD group).

Contrasting non-ASD and ASD on COS. A Mann–Whitney test indicated that the ASD group (*Mdn* = 31) scored significantly lower on COS-Researcher than the non-ASD group (*Mdn* = 37.5), $U = 413.50$, $z = -4.01$, $p < 0.001$. However, median scores on COS-Teacher were not significantly different between the ASD group (*Mdn* = 23) and the non-ASD group (*Mdn* = 25), $U = 502.00$, $z = -1.55$, $p = 0.12$. Figure 2 shows the relative frequency distribution of scores on COS-Teacher (top panel) and COS-Researcher (bottom panel) for the ASD and non-ASD groups.

As a potential alternative to the original 5-point scoring system of COS (Au et al., 2021), we further transformed the COS rating into binary scoring: item scores of 1 and 2 were converted to 0, signifying problem in the observed behavior; scores of 3, 4, and 5 were converted to 1. This procedure was initiated based on our observation that some teachers in this study found it hard to differentiate between occurrence of behavior that was "rare" versus "less often than most students," or between "more often" versus "much more often than most students" while rating on COS. Based on the transformed scores, the ASD group's median scores (2 and 6, respectively) were significantly lower than the non-ASD group's (3 and 10, respectively) on both COS-Teacher ($U = 432.50$, $z = -2.26$, $p = 0.02$) and COS-Researcher ($U = 428.00$, $z = -3.92$, $p < 0.001$) using Mann–Whitney tests.

Table 1. Confusion matrices for the performance of the two screening approaches in identifying preschoolers with ASD.

Clinical assessment	Approach 1[a]		Approach 2[b]	
	Screen-positive	Screen-negative	Screen-positive	Screen-negative
ASD (*n* = 18)	13	5	15	3
Non-ASD (*n* = 112)	51	61	41	71
Total (*n* = 130)	64	66	56	74

ASD: autism spectrum disorder; COS: Classroom Observation Scale.
[a]At or below the 15th percentile on COS-Teacher and below the median on COS-Researcher.
[b]At or below the 15th percentile on COS-Researcher and below the median on COS-Teacher.

Table 2. Validity of the two screening approaches in identifying preschoolers with ASD.

	χ^2 (*p*-value)	Cramer's *V* (*p*-value)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR− (95% CI)	OR (95% CI)	z statistic for OR (*p*-value)
Approach 1[a]	4.43 (0.04)	0.18 (0.04)	0.72 (0.47–0.90)	0.54 (0.45–0.64)	1.59 (1.12–2.25)	0.51 (0.24–1.10)	3.11 (1.04–9.31)	2.03 (0.04)
Approach 2[b]	13.82 (< 0.001)	0.33 (< 0.001)	0.83 (0.59–0.96)	0.63 (0.54–0.72)	2.28 (1.65–3.13)	0.26 (0.09–0.75)	8.66 (2.36–31.70)	3.26 (0.001)

ASD: autism spectrum disorder; CI: confidence interval; LR: likelihood ratio; OR: odds ratio; COS: Classroom Observation Scale.
LR+: sensitivity/(1−specificity); LR−: (1−sensitivity)/specificity; OR: LR+/LR−: (sensitivity × specificity)/((1−sensitivity)(1−specificity)).
[a]At or below the 15th percentile on COS-Teacher and below the median on COS-Researcher.
[b]At or below the 15th percentile on COS-Researcher and below the median on COS-Teacher.

Identifying preschoolers with ASD near the end of second year. Table 1 shows the confusion matrices for the performance of the two screening approaches in identifying preschoolers with ASD diagnosed about 1 year after the screening. We estimated the sensitivity, specificity, and odds ratio (OR) for the two screening approaches (Table 2) using Pearson's chi-square tests to assess relations between COS screening and subsequent ASD diagnoses. Both screening approaches (Figure 1) significantly predicted ASD versus non-ASD classification (COS-Teacher: χ^2 = 4.43, *p* = 0.04; COS-Researcher: χ^2 = 13.82, *p* < 0.001), and the effect size (Cramer's *V*) was near medium for COS-Researcher and small but nonetheless significant for COS-Teacher. OR values for both screening approaches were significantly greater than 1. The OR using the bottom 15% cutoff was higher for COS-Researcher (OR = 8.66, 95% confidence interval (CI) = (2.36, 31.70), *p* = 0.001) than COS-Teacher (OR = 3.11, 95% CI = (1.04, 9.31), *p* = 0.04), indicating better screening accuracy of the former. For the randomly selected screen-negative peers (*n* = 55), only one met diagnostic criteria for ASD.

Separate receiver operating characteristic (ROC) analyses further assessed the validity of COS-Teacher and COS-Researcher in discriminating ASD versus non-ASD cases (Figure 3), where larger area under the ROC curve (AUC) indicates better screening accuracy, and an AUC above 0.7 shows at least moderate accuracy. The AUCs for COS-Teacher and COS-Researcher were 0.62 (*p* = 0.12) and 0.80 (*p* < 0.001), respectively, based on the original 5-point scoring system. For the transformed binary scores, the AUCs were 0.67 (*p* = 0.03) and 0.79 (*p* < 0.001), respectively, for COS-Teacher and COS-Researcher. Hence, both COS-Teacher and COS-Researcher helped to identify children who were more likely than their peers to have ASD by predicting ASD diagnosis about 1 year later above chance (AUC > 0.5). The screening validity was apparently higher for COS-Researcher than COS-Teacher.

Discussion

As in Au et al. (2021), COS helped observers with little or no clinical training—research assistants with just a few hours of COS training and preschool teachers with an hour of briefing—to identify children in their first year of Chinese-language preschool as more likely than their peers to have ASD. Reliability estimates of COS-Teacher and COS-Researcher in this study were comparable to those for the original English COS. Validity of COS-Researcher here closely aligned with prior results. For both the English COS (Au et al., 2021) and Chinese COS here, the Cramer's *V* measuring the strength of association between the COS-Researcher classification status and subsequent ASD diagnosis was 0.33, with sensitivity and specificity, respectively, around 0.80 and 0.65. Moreover, the AUCs for COS-Researcher in both studies were around 0.80 (*p*-values < 0.001).

However, although predictive ORs for COS-Teacher were significant in both studies, the OR value was apparently lower for the Chinese COS (OR = 3.11) than for the English COS (OR = 14.63). The disparity in
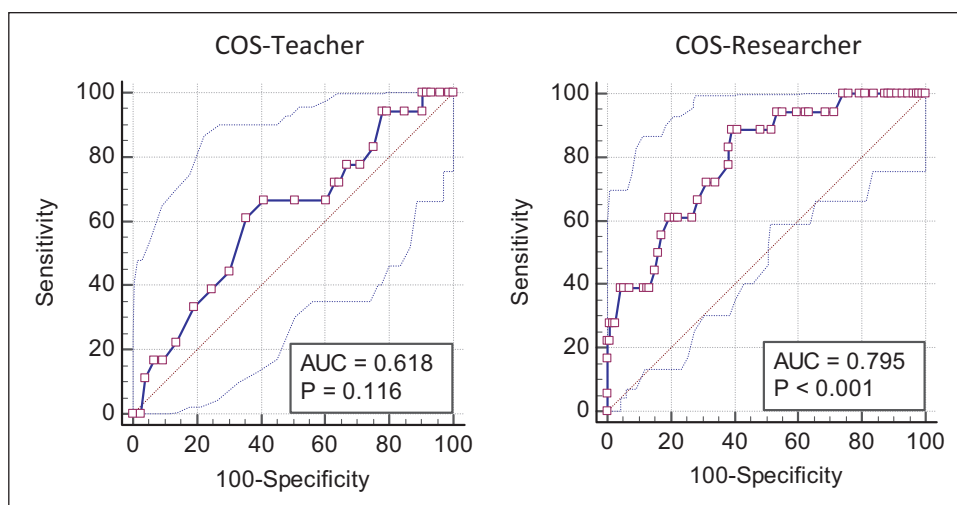
Figure 3. Receiver operating characteristic (ROC) curves for COS-Teacher (left) and COS-Researcher (right) in predicting ASD diagnosis based on ADOS-2 (supplemented by ADI-R). Screening accuracy was measured by the area under the ROC curve (AUC).

teacher–student ratios between the international preschools and less-resourced Chinese-language preschools might have contributed to the results. Note also that the research assistants had received 6 h of training on using COS, whereas the teachers only received about an hour of training. Perhaps the teachers in the Chinese-language preschools needed more training than the better-resourced international preschool teachers, who might be more aware of ASD symptomatology to begin with. Future studies should explore whether screening accuracy of the teacher-report on COS can be enhanced by more intensive training for the preschool teachers, especially where staff development opportunities and teacher–child ratios are less favorable. Moreover, possible cultural differences in symptom endorsement and differences in the educational background of teachers between the international versus Chinese-language preschools might also have contributed to the discrepant ORs for COS-Teacher in Au et al. (2021) and this study. Due to cultural differences in expectations on child's developmental behavior (Matson et al., 2011), teachers at English-speaking international preschools in Hong Kong—mostly expatriates from North America, Europe, and Australia—may hold slightly different criteria in contrast to locally trained Hong Kong Chinese preschool teachers when judging whether certain behaviors are problematic. These factors can be further explored in future studies. Finally, the higher screening accuracy on ROC analyses for COS-Teacher when scoring was binary relative to finer-grained hinted at the potential benefit of using a binary scoring system for COS, which deserves more investigation.

To conclude, this study provided further evidence on the versatility and ecological validity of the COS for use by preschool teachers and non-clinically trained observers in the early identification of children with ASD in community settings.

ORCID iD

Kathy Kar-man Shum [iD] https://orcid.org/0000-0003-4340-3160

References

Achenbach, T. M. (2014). *DSM-oriented guide for the Achenbach System of Empirically Based Assessment (ASEBA)*. Research Center for Children, Youth, and Families, The University of Vermont.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles* (Vol. 30). Research Center for Children, Youth, and Families, The University of Vermont.

Au, A. H. C., Shum, K. K. M., Cheng, Y. T., Tse, M. Y., Wong, R. M. F., Li, J., & Au, T. K. (2021). Autism spectrum disorder screening in preschools. *Autism*, *25*(2), 516–528.

Corbett, B. A., Schupp, C. W., Simon, D., Ryan, N., & Mendoza, S. (2010). Elevated cortisol during play is associated with age and social engagement in children with autism. *Molecular Autism*, *1*, Article 13.

Education Bureau, Hong Kong SAR. (2020). *Overview of kindergarten education in Hong Kong*. https://www.edb.gov.hk/

en/edu-system/preprimary-kindergarten/overview/index.html

Kim, S., & Lord, C. (2012). New autism diagnostic interview-revised algorithms for toddlers and young preschoolers from 12-47 months of age. *Journal of Autism and Developmental Disorders*, *42*(1), 82–93.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Li, H., Rao, N., & Tse, S. K. (2012). Adapting Western pedagogies for Chinese literacy instruction: Case studies of Hong Kong, Shenzhen, and Singapore preschools. *Early Education and Development*, *23*(4), 603–621.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) manual (part 1): Modules 1–4*. Western Psychological Services.

Matson, J. L., Worley, J. A., Fodstad, J. C., Chung, K. M., Suh, D., Jhin, H. K., . . .Furniss, F. (2011). A multinational study examining the cross cultural differences in reported symptoms of autism spectrum disorders: Israel, South Korea, the United Kingdom, and the United States of America. *Research in Autism Spectrum Disorders*, *5*(4), 1598–1604.

Oosterling, I., Roos, S., Bildt, A., Rommelse, N., Jonge, M., Visser, J., . . .Buitelaar, J. (2010). Improved diagnostic validity of the ADOS revised algorithms: A replication study in an independent sample. *Journal of Autism and Developmental Disorders*, *40*, 689–703.

Peach, M. (1994). The nonacademic curriculum of the Japanese preschool. *Childhood Education*, *71*(1), 9–13.

Rescorla, L. A., Ghassabian, A., Ivanova, M. Y., Jaddoe, V. W., Verhulst, F. C., & Tiemeier, H. (2019). Structure, longitudinal invariance, and stability of the Child Behavior Checklist 1½-5's *Diagnostic and Statistical Manual of Mental Disorders*-Autism Spectrum Disorder scale: Findings from Generation R (Rotterdam). *Autism*, *23*(1), 223–235.

Rescorla, L. A., Given, C., Glynn, S., Ivanova, M. Y., & Achenbach, T. M. (2019). International comparisons of autism spectrum disorder behaviors in preschoolers rated by parents and caregivers/teachers. *Autism*, *23*(8), 2043–2054.