International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Behavioral Domains Associated with Autism in Toddlers: A Secondary Analysis of the Polish Q-CHAT Dataset

Ruslan Kurmashev

Abstract: ***Total scores on the Quantitative Checklist for Autism in Toddlers (Q-CHAT) may conceal differences among behavioral domains. This secondary cross-sectional analysis compared Social/Communication, Non-social/Behavioral, and Speech/Language scores in an open Polish dataset of 252 toddlers aged 18-24 months, including 135 records coded as ASD and 117 controls. Domain separation was quantified using Hedges' g with bootstrap 95% confidence intervals; age- and sex-adjusted logistic models were fitted separately and jointly. Social/Communication showed the largest separation (g = 1.94, 95% CI 1.65-2.26), compared with Non-social/Behavioral (g = 0.47) and Speech/Language (g = 0.69). In the joint model, Social/Communication retained a strong association with ASD status (OR per pooled SD 18.43, 95% CI 8.66-39.22), whereas the other domain confidence intervals included 1. Speech/Language internal consistency was low (alpha = 0.20). Q-CHAT group differences were therefore strongly domain-dependent and concentrated in Social/Communication in this dataset. These findings do not validate the imported factor structure, a diagnostic cutoff, standalone diagnostic use, or population screening performance.***

Keywords: autism; toddlers; Q-CHAT; behavioral domains; secondary analysis

1. Introduction

Behavioral characteristics associated with autism can be observable during toddlerhood, but caregiver questionnaires are most defensible as tools for trait measurement and developmental surveillance rather than as standalone diagnostic tests. The 25-item Quantitative Checklist for Autism in Toddlers (Q-CHAT) was developed to measure autistic traits dimensionally at 18-24 months, although its initial report described clinical validation as incomplete [1]. A systematic review subsequently characterized Q-CHAT as promising while emphasizing the need for further validation [2]. Prospective population research also showed that cases may emerge across repeated assessments and that positive predictive value is substantially lower than separation observed in selected case-control samples [3].

A total score can obscure whether group differences are broadly distributed or concentrated in particular behaviors. Exploratory analysis in a Singaporean community cohort grouped Q-CHAT items into Social/Communication, Non-social/Behavioral, and Speech/Language domains [4]. This structure offers a published basis for domain-level comparison, but factor membership derived in one cultural and sampling context cannot be assumed to represent identical constructs elsewhere. Meaningful cross-group interpretation requires evidence that measurement structure is sufficiently comparable across populations [5].

An open Polish dataset provides item-level Q-CHAT responses for toddlers labelled as having autism spectrum disorder (ASD) and typically developing controls [6]. A separate, larger Polish study concluded that further adaptation and validation of the translated questionnaire were required [7]. The open dataset has been used for machine-learning classification [8], but such analyses do not establish its diagnostic provenance or clarify how much the three published behavioral domains contribute to ASD-control separation. A conventional domain-level analysis can address this narrower descriptive question while avoiding claims of diagnostic validation or prospective prediction.

This study therefore examined whether ASD-control differences were distributed similarly across the three published Q-CHAT domains. The primary objectives were to estimate standardized group separation for each domain and directly compare the Social/Communication effect size with the other two effects. Secondary analyses estimated age- and sex-adjusted associations for each standardized domain separately and jointly. We expected higher domain scores in the ASD group, the largest separation for Social/Communication, and differing adjusted association strengths across domains. Item-level heterogeneity was evaluated exploratorily.

2. Literature Survey

The original Quantitative Checklist for Autism in Toddlers (Q-CHAT) was developed as a 25-item caregiver questionnaire with ordinal responses intended to measure autistic traits dimensionally at 18-24 months. The preliminary samples differed in score and age, and clinical validation remained incomplete [1]. A 10-item form derived from retrospective case-control data showed high sensitivity and specificity, but differing administration procedures and absent independent diagnostic verification restricted prospective interpretation [9]. These studies established that Q-CHAT scores can differentiate selected groups, not that either form functions as a standalone diagnostic instrument. Consistent with that distinction, a systematic review of 52 reports covering 16 toddler screening measures characterized Q-CHAT as promising while concluding that further validation was required [2].

Evidence also indicates that Q-CHAT responses may not be adequately represented by a total score alone. In a Singaporean community cohort, exploratory factor analysis yielded Social/Communication, Non-social/Behavioral, and Speech/Language factors comprising 10, 8, and 4 items;

together they explained 38.1% of the variance, and reliability was weakest for the total and Speech/Language scores [4]. An Italian clinical study differentiated autistic children from children with developmental delay and typically developing children, but age, group-size, and sampling differences limited cutoff transportability [10]. Serbian data supported a different 17-item three-factor solution, although only 26 autistic toddlers were included and diagnosis was not established with a standardized instrument [11]. Meaningful comparisons require evidence that constructs are represented comparably across populations [5]. Accordingly, use of the Singaporean item grouping in Polish data is a hypothesis-driven secondary scoring strategy; it does not confirm the original factor structure or establish measurement invariance.

Prospective evidence shows that apparent performance depends strongly on setting and design. In a population study, 3,770 of 13,070 invited families returned the Q-CHAT and 121 children underwent phase-one assessment. Eleven autism cases were identified; positive predictive value was 17% at a cutoff of 39 initially and 28% after follow-up, while additional cases emerged later [3]. In a community pediatric sample, ordinal Q-CHAT-10 scoring improved sensitivity relative to some alternatives, but specificity and participation-related generalizability remained trade-offs [12]. Diagnostic-accuracy estimates can also be inflated by nonrepresentative participants or comparisons of clinically selected cases with convenient controls [13]. Therefore, discrimination in an ASD-control dataset, including an in-sample area under the receiver operating characteristic curve, cannot estimate population screening performance, predictive value, diagnostic validity, or clinical utility.

The available Polish evidence leaves a specific domain-level question unresolved. The open 2020 dataset contains Q-CHAT responses for 252 toddlers described as autistic or typically developing, but its stated group counts sum to 253 [6]. A later Polish study analyzed a different sample of 1,024 toddlers defined largely by parent-reported concerns or developmental-risk categories; eight children with confirmed autism were excluded from the principal analyses. Its authors concluded that further cultural adaptation and validation of the Polish Q-CHAT were needed [7]. Caregiver scores also require caution because child language, maternal depressive symptoms, and maternal education were associated with particular domains in Singaporean data [14], while language and ethnicity were associated with total scores among children born very preterm [15]. A machine-learning analysis of the public Polish dataset did not establish recruitment, diagnostic adjudication, clinical validity, or the relative contribution of the three domains [8]. Within the reviewed pre-cutoff literature, no study compared these domains in this exact dataset using standardized group contrasts and age- and sex-adjusted associations. This narrower gap defines the present problem without presuming diagnostic performance or validation.

3. Problem Definition

The unresolved problem is whether ASD-control differences in the Polish Q-CHAT dataset are distributed similarly across the Social/Communication, Non-social/Behavioral, and Speech/Language domains reported in the Singaporean factor study [4,6]. Applying those published item groups without re-estimating factors after inspecting ASD status avoids outcome-driven construction of the domain scores. It nevertheless assumes that the item grouping is sufficiently transportable for a secondary comparison and does not establish the factor structure or measurement invariance in Polish data [5]. The intended estimands are therefore standardized group separation and sample-specific adjusted associations. They are not diagnostic accuracy, population screening performance, future risk, causal effects, or specificity relative to other developmental conditions.

The primary objective was to estimate ASD-control separation within each domain and compare the Social/Communication standardized mean difference with those of the other two domains. The secondary objective was to estimate the association of each standardized domain with ASD status after adjustment for age and sex, both separately and jointly. We hypothesized that scores would be higher in the ASD group across all three domains, that separation would be greatest for Social/Communication, and that adjusted association strengths would differ across domains. Heterogeneity across the 25 individual items was examined as an exploratory objective. Because no dated analysis plan predating access to the results was available, the analyses are described as primary or exploratory rather than confirmatory.

4. Methodology/Approach

4.1 Study Design and Data Source

We conducted a secondary cross-sectional analysis of the openly available dataset *Q-CHAT scores of Polish toddlers with autism spectrum disorders and typically developing controls*, version 2, released on 12 October 2020 under a CC BY 4.0 license [6]. Reporting was guided by the Strengthening the Reporting of Observational Studies in Epidemiology statement [16]. The source SPSS file contained 252 unique child identifiers: 135 records coded as ASD and 117 coded as control, with ages from 18 to 24 months. The repository description reports 118 controls although its stated total is 252; all analyses used the 117 control records present in the source file. The repository characterizes ASD status as a community-based diagnosis. The available documentation did not identify recruitment dates, eligibility criteria, diagnostic instruments, assessors, adjudication procedures, the originating ethics committee, approval number, or consent process. We therefore did not infer these details. The present analysis used only de-identified open data and involved no participant contact.

4.2 Q-CHAT Domain Scores

The released dataset contains 25 caregiver-reported Q-CHAT items scored from 0 to 4, with higher recoded values indicating more atypical behavior. We applied the three-

factor membership reported by Magiati et al. [4] without re-estimating factors in the outcome-labelled sample. The Social/Communication score was the unweighted sum of Q1, Q2, Q5, Q6, Q9, Q10, Q12, Q15, Q19, and Q21. The Non-social/Behavioral score summed Q7, Q11, Q16, Q20, Q22, Q23, Q24, and Q25. The Speech/Language score summed Q4, Q8, Q17, and Q18. Q3, Q13, and Q14 were not included in these domain scores. A domain score was calculated only when all constituent items were available. Domain scores were treated as secondary composites derived from a published exploratory structure, not as psychometrically confirmed Polish subscales.

4.3 Data Quality and Missing Data

SPSS-declared user-missing values were converted to missing before analysis. We verified row counts, group counts, identifier uniqueness, item ranges, and the correspondence between the released total and the raw item sum. One control record, child identifier `bdbp0221`, contained Q4 = 11 although valid item values were 0-4. Because the intended response could not be recovered, only this value was set to missing in the primary analysis; the participant remained in analyses not requiring Q4. Social/Communication and Non-social/Behavioral analyses therefore included all 252 participants, whereas Speech/Language and the cleaned 25-item total included 251. No values were imputed. Each regression used complete observations for its specified outcome and predictors. As a sensitivity analysis, the invalid Q4 value was replaced with 4, the maximum valid score, without interpreting 4 as the participant's true response.

4.4 Statistical Analysis

Continuous variables were summarized using means and standard deviations, with medians added for domain scores; categorical variables were summarized as counts and percentages. Descriptive ASD-control comparisons in the demographic table used Welch's two-sample t test for continuous variables and Pearson's chi-square test for categorical variables. Domain distributions were additionally compared with Welch's test and the two-sided Mann-Whitney U test, but effect estimates and confidence intervals were emphasized.

For each domain, ASD-control separation was expressed as Hedges' g, calculated as the bias-corrected standardized mean difference for ASD minus control [17]. Percentile 95% confidence intervals were obtained from 5,000 bootstrap resamples with fixed pseudorandom seeds. Differences between the Social/Communication effect size and each of the other two domain effect sizes were bootstrapped by resampling participants within the ASD and control groups while preserving paired domain scores within each participant.

Domain scores were standardized using the mean and sample standard deviation of the pooled observed sample. Binary logistic regression modelled ASD status as the outcome. Separate models included one standardized domain, age in months, and sex (male versus female); the joint model included all three standardized domains, age, and sex. Associations are reported as odds ratios per one pooled standard deviation with 95% Wald confidence intervals. Variance inflation factors were calculated for the joint model. Because the dataset was ASD-enriched and did not arise from population screening, odds ratios were interpreted only as sample-specific associations and not as prevalence, risk, or prospective prediction estimates. Two-sided p values below 0.05 were treated as statistical evidence, but no multiplicity adjustment was applied to the three domain-level models; estimates and confidence intervals remained primary.

Internal consistency was summarized using Cronbach's alpha in the full sample and separately within the ASD and control groups, together with corrected item-total correlations. Alpha was interpreted descriptively and was not treated as evidence of unidimensionality or factor validity [18]. Exploratory age- and sex-adjusted logistic models were then fitted for each of the 25 ordinal item scores, entered as a linear one-point predictor. This parameterization assumes a constant log-odds change between adjacent response categories. The Benjamini-Hochberg procedure controlled the false discovery rate across the 25 item-level tests [19].

Exploratory nested logistic models were fitted to the same complete-case sample to compare Social/Communication plus age and sex with models that added one or both remaining domains. Model fit was summarized using the likelihood-ratio test, Akaike and Bayesian information criteria, McFadden's pseudo-R^2, and the area under the receiver operating characteristic curve. All areas under the curve were apparent, in-sample estimates without optimism correction or external validation. Sensitivity analyses added prematurity to the joint model, restricted the analysis to term-born participants, used heteroscedasticity-consistent HC3 standard errors, replaced the invalid Q4 value with 4, and tested domain-by-sex interactions. Analyses were implemented in Python using NumPy, pandas, SciPy, statsmodels, scikit-learn, and Matplotlib; the reproducible script and machine-readable outputs were preserved in the project handoff package.

5. Results & Discussion

5.1 Sample and Data Quality

The source file contained 252 participants, including 135 in the ASD group and 117 controls; all child identifiers were unique. The ASD group was younger on average than the control group and contained a higher proportion of boys (Table 1). Preterm birth, birth weight, and maternal education did not show clear group differences. Birth weight was available for 240 participants, and maternal education was available for 241. The single invalid Q4 value occurred in a control record. After it was set to missing, Speech/Language and the cleaned total-score analyses included 251 participants, whereas the other two domains included all 252.

Table 1: Participant characteristics by released group label

Characteristic	ASD (n = 135)	Control (n = 117)	p
Age, months, mean (SD)	20.76 (2.14)	21.42 (1.97)	0.011
Male sex, n (%)	96 (71.1)	61 (52.1)	0.002
Preterm birth, n (%)	20 (14.8)	10 (8.5)	0.125
Birth weight, g, mean (SD)	3248.8 (740.8), n = 128	3384.2 (534.4), n = 112	0.103
Maternal education: primary/vocational, n (%)	8 (6.2)	7 (6.2)	0.942†
Maternal education: high school, n (%)	33 (25.8)	27 (23.9)	—
Maternal education: higher, n (%)	87 (68.0)	79 (69.9)	—

Values are based on available observations. Continuous variables were compared using Welch's test and categorical variables using Pearson's chi-square test. †Overall comparison across the three maternal-education categories.

5.2 Domain-Level Group Differences

Mean scores were higher in the ASD group for every domain (Table 2). Social/Communication showed the largest standardized difference, with a mean of 20.41 (SD 7.23) in the ASD group and 8.38 (SD 4.73) in controls (Hedges' g = 1.94, 95% CI 1.65-2.26). The Non-social/Behavioral difference was smaller (g = 0.47, 95% CI 0.23-0.71), while Speech/Language was intermediate (g = 0.69, 95% CI 0.45-0.95). Welch and Mann-Whitney tests yielded $p < 0.001$ for all three comparisons. The Social/Communication effect exceeded the Non-social/Behavioral effect by 1.47 standardized units (bootstrap 95% CI 1.13-1.82) and the Speech/Language effect by 1.25 units (95% CI 0.90-1.61). These standardized differences are summarized in Figure 1.

Table 2: Domain scores and adjusted associations with ASD status

Domain	ASD mean (SD)	Control mean (SD)	Separate adjusted OR (95% CI)	Joint adjusted OR (95% CI)
Social/Communication	20.41 (7.23), n = 135	8.38 (4.73), n = 117	18.91 (9.05-39.54)	18.43 (8.66-39.22)
Non-social/Behavioral	9.33 (4.81), n = 135	7.20 (4.15), n = 117	1.68 (1.27-2.22)	0.83 (0.53-1.32)
Speech/Language	7.73 (3.00), n = 135	5.78 (2.63), n = 116	2.03 (1.51-2.75)	1.24 (0.81-1.91)

ORs are reported per one pooled-sample SD and are adjusted for age and sex. Separate models contain one domain; the joint model contains all three domains. Joint-model n = 251. CI, confidence interval; OR, odds ratio.

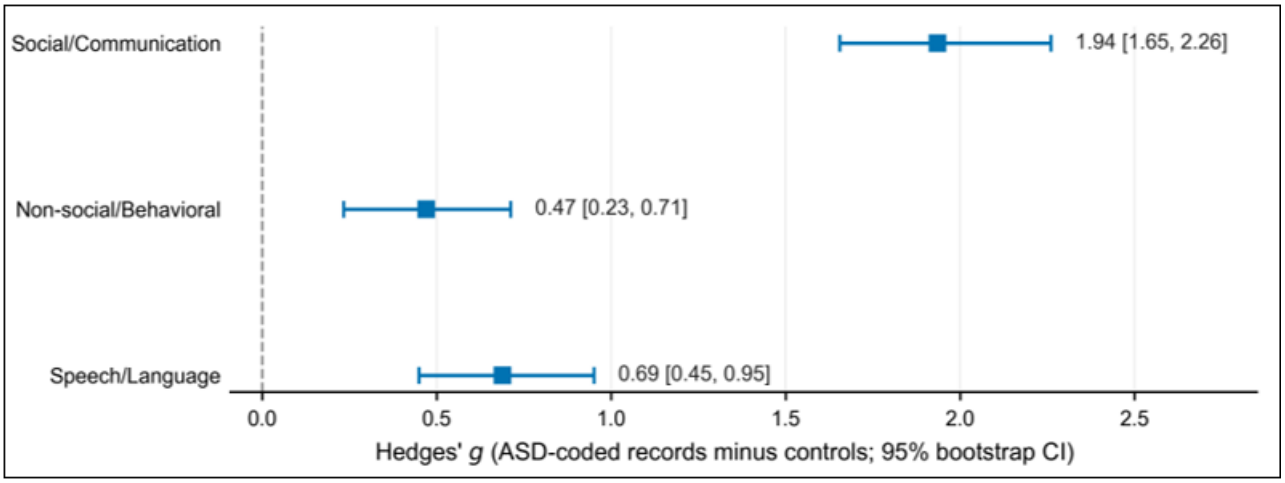

Figure 1: Standardized ASD-control differences across prespecified Q-CHAT domains. Squares represent Hedges' g for ASD-coded records minus controls, and horizontal lines represent percentile 95% confidence intervals obtained from 5,000 within-group bootstrap resamples. Positive estimates indicate higher domain scores in ASD-coded records. Social/Communication and Non-social/Behavioral analyses included 135 ASD-coded and 117 control records; Speech/Language included 135 and 116 records, respectively, because one invalid Q4 value was treated as missing.

5.3 Adjusted and Exploratory Models

In separate age- and sex-adjusted models, each pooled-SD increase was associated with higher odds of membership in the ASD group: OR 18.91 (95% CI 9.05-39.54) for Social/Communication, 1.68 (95% CI 1.27-2.22) for Non-social/Behavioral, and 2.03 (95% CI 1.51-2.75) for Speech/Language. When all three domains were entered jointly, Social/Communication remained strongly associated with group status (OR 18.43, 95% CI 8.66-39.22; $p < 0.001$). The conditional estimates for Non-social/Behavioral (OR 0.83, 95% CI 0.53-1.32; p = 0.434) and

Speech/Language (OR 1.24, 95% CI 0.81-1.91; $p = 0.327$) were imprecise and their confidence intervals included 1. Variance inflation factors ranged from 1.03 to 1.46.

The apparent area under the curve was 0.928 for the Social/Communication, age, and sex model and 0.929 for the full-domain model. Adding both remaining domains did not improve fit in the likelihood-ratio comparison, $\chi^2(2) = 1.39$, $p = 0.499$; AIC and BIC were also lower for the Social/Communication model. These estimates describe only in-sample discrimination.

Internal consistency differed markedly by domain. Social/Communication alpha was 0.86 overall, 0.76 in the ASD group, and 0.71 in controls. Corresponding values were 0.60, 0.55, and 0.65 for Non-social/Behavioral and 0.20, 0.18, and 0.27 for Speech/Language. Q18 had a negative corrected item-total correlation within Speech/Language ($r = -0.32$), while Q22 was approximately uncorrelated with the remaining Non-social/Behavioral items ($r = -0.01$). Sixteen of the 25 item-level associations remained below the Benjamini-Hochberg false-discovery-rate threshold. The largest positive adjusted item association was observed for eye contact (OR per point 10.27, 95% CI 5.66-18.62), whereas Q18 and Q22 showed inverse associations. The complete set of item-level estimates is shown in Figure 2.

Sensitivity analyses retained the dominant Social/Communication association. Its joint-model OR was 18.43 after adjustment for prematurity, 14.33 (95% CI 6.86-29.94) among term-born participants, and 18.72 (95% CI 8.79-39.87) when the invalid Q4 value was replaced with 4. HC3 standard errors widened its confidence interval to 7.76-43.80 without changing the point estimate. Domain-by-sex interaction p values ranged from 0.108 to 0.991 and did not provide clear evidence of effect modification.

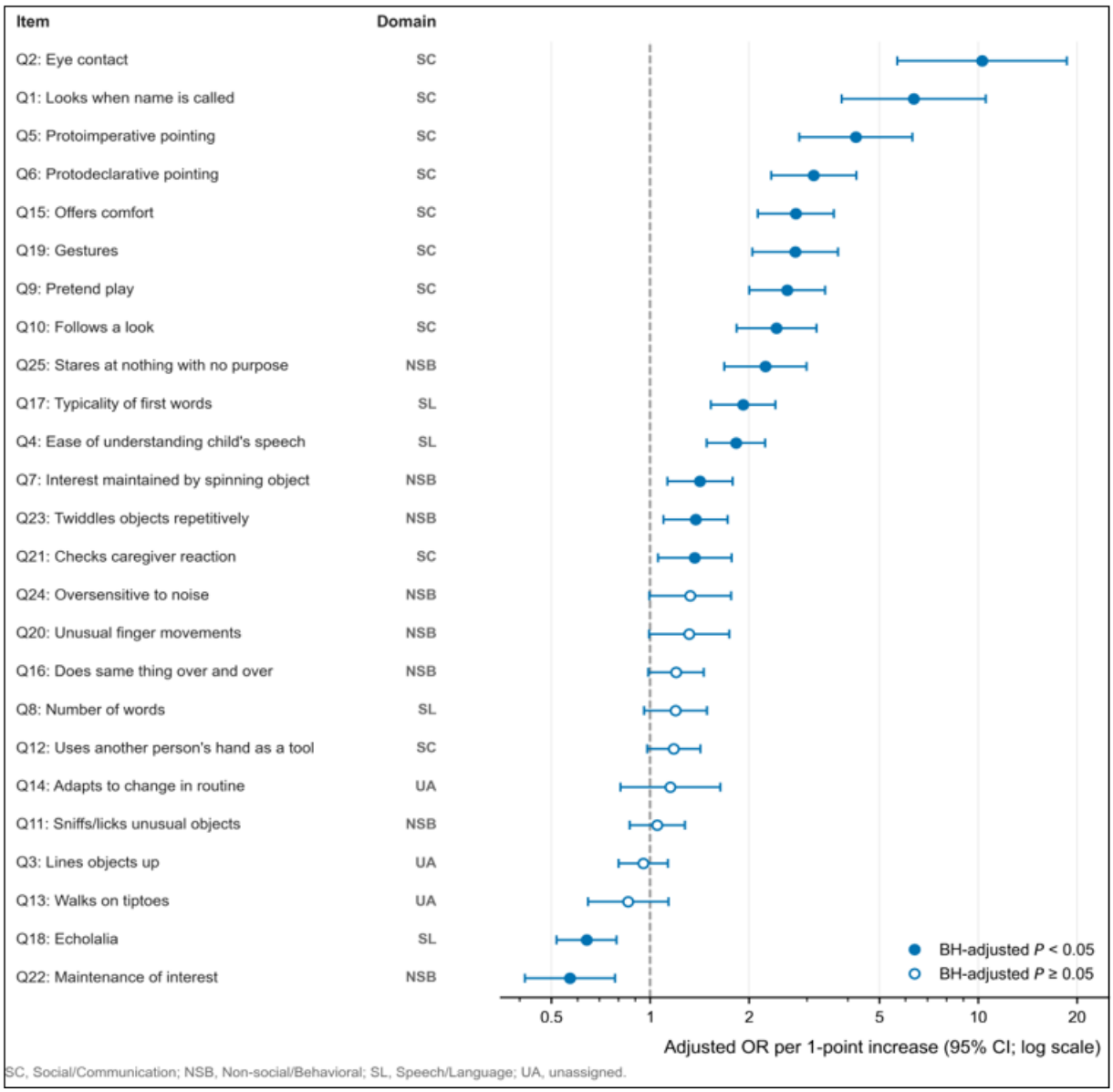

Figure 2: Exploratory item-level associations with ASD versus control group membership. Points represent odds ratios from separate logistic regression models for a one-point increase in the coded Q-CHAT item score, adjusted for age and sex; horizontal lines represent 95% Wald confidence intervals. Items are ordered by the adjusted odds-ratio point estimate. Filled markers indicate Benjamini-Hochberg-adjusted $P < 0.05$ across 25 item-level tests. Models included 252 records, except Q4, for which one invalid value was treated as missing ($n = 251$). Items were entered as linear ordinal predictors, so estimates assume a constant log-odds change between adjacent response categories. The estimates should not be interpreted as diagnostic performance, protective effects, or comparative item importance. SC, Social/Communication; NSB, Non-social/Behavioral; SL, Speech/Language; UA, unassigned. Q3, Q13, and Q14 were not assigned to the three imported domains.

5.4 Principal Interpretation

The findings indicate that ASD-control separation was strongly domain-dependent in this dataset. The data were consistent with higher scores across all three domains and with the prespecified expectation that Social/ Communication would show the greatest separation. The bootstrap contrasts quantify this difference directly rather than inferring it from separate significance tests. Social/Communication also retained a similar association after conditioning on the other domains, age, and sex, whereas the other domains did not show clear conditional associations. This pattern describes the released sample and scoring structure; it does not imply that autism is exclusively or primarily social-communicative.

The weaker conditional associations for Non-social/ Behavioral and Speech/Language should not be interpreted as evidence that these characteristics are absent or unimportant. Both domains differed between groups when considered separately. Their attenuation in the joint model may reflect information shared with Social/Communication, heterogeneous item behavior, and unequal measurement precision. Low VIFs argue against severe numerical multicollinearity, but they do not make the conditional coefficients equivalent to distinct causal effects. The very low Speech/Language alpha and inverse behavior of Q18 are particularly important because alpha cannot establish unidimensionality [18]. Magiati et al. likewise reported weaker reliability for Speech/Language [4], while the different Serbian item structure illustrates that domain composition may vary across settings [11].

5.5 Relation to Earlier Evidence and Clinical Boundaries

Applying the Singapore-derived domains to Polish data allowed a transparent hypothesis-driven comparison, but did not demonstrate cross-cultural measurement equivalence [4,5]. The larger Polish study called for further adaptation and validation of the translated Q-CHAT [7]. Domain scores may also reflect factors beyond autism status: language and caregiver characteristics were associated with Q-CHAT domains in Singapore [14], and language and ethnicity were associated with total scores among children born preterm [15]. The persistence of the Social/Communication association after prematurity adjustment and in term-born participants supports robustness within this sample, but it cannot establish specificity or mechanism.

The high apparent AUC must be separated from prospective screening performance. Population Q-CHAT research reported substantially lower positive predictive values and identified additional cases at later follow-up [3]. More generally, diagnostic estimates can be inflated when clinically selected cases are compared with nonrepresentative controls [13]. The present dataset has no developmental-delay or other clinical comparison group, unlike the Italian clinical study [10], and therefore cannot establish autism-specific discrimination. The ROC results do not justify a cutoff, diagnostic claim, or recommendation for population screening.

5.6 Strengths and Limitations

Strengths include use of an openly accessible dataset, preservation of the original SPSS file, explicit auditing of identifiers, declared missing values, and item ranges, and transparent handling of the invalid Q4 value. Domain membership was taken from a pre-existing publication rather than constructed after examining group status. The analysis emphasized standardized effects with bootstrap confidence intervals, compared effects using participant-preserving resampling, adjusted for age and sex, controlled false discovery rate for exploratory item tests, and examined several sensitivity specifications. Code and machine-readable outputs were retained.

Several limitations constrain interpretation. This was a cross-sectional secondary analysis of an ASD-enriched comparison sample; temporal order, causal mechanisms, prevalence, calibration, predictive values, and population screening performance cannot be inferred. The repository does not document recruitment, eligibility criteria, formal diagnostic adjudication, or the original ethics and consent procedures. The absence of a developmental-delay group prevents evaluation of autism specificity, and caregiver reporting may introduce common-method or reporting bias. The Polish translation and the imported three-factor structure were not validated in this sample, and no measurement-invariance analysis was performed. Domain lengths and reliability differed, one item value was invalid, and complete-case handling was not a general missing-data model. Item scores were entered linearly in exploratory logistic models, model diagnostics beyond VIF and robust-standard-error sensitivity were limited, and sex-interaction analyses may have had low power. Finally, no dated analysis plan predating access to results was available; the hypotheses should therefore not be interpreted as prospectively preregistered.

6. Conclusion

This secondary analysis showed that ASD-control differences in the Polish Q-CHAT dataset were not evenly distributed across the three published behavioral domains. Social/Communication produced the largest standardized

separation (Hedges' g = 1.94) and exceeded the Non-social/Behavioral and Speech/Language effects by 1.47 and 1.25 standardized units, respectively. It also retained a strong age- and sex-adjusted association after all three domains were entered jointly. Non-social/ Behavioral and Speech/ Language scores were higher in the ASD group when analyzed separately, but their conditional estimates in the joint model were less precise and their confidence intervals included 1.

These results support examining Q-CHAT responses at domain and item levels rather than relying exclusively on the total score. However, the very low internal consistency of Speech/Language, heterogeneous item behavior, and uncertain cross-cultural transportability limit interpretation of the composite scores. The contribution of this study is therefore a reproducible domain-level characterization of one openly available ASD-control dataset. It does not confirm the three-factor structure in Polish toddlers, establish causation or autism specificity, validate a diagnostic threshold, or estimate prospective screening performance. Social/Communication should consequently be understood as the dominant source of group separation in this sample, not as a standalone diagnostic measure or a complete description of early autism-related behavior.

7. Future Scope

Future work should first clarify the dataset's recruitment and diagnostic provenance and independently evaluate the Polish Q-CHAT translation. Confirmatory factor analysis and measurement-invariance testing are needed in newly recruited Polish cohorts with standardized diagnostic assessment and comparison groups that include developmental delay and other clinically relevant conditions. Prospective repeated-measure studies should examine score stability, developmental change, and external validity before domain scores are used for screening or prediction. The inverse behavior of Q18 and weak coherence of Q22 should be investigated as prespecified psychometric questions rather than treated as isolated post hoc findings. Adequately powered studies should also evaluate whether domain associations differ by sex, prematurity, language level, or caregiver characteristics. Any future prediction model should use representative sampling, independent validation, calibration assessment, and explicit optimism correction. Until such evidence is available, these domain scores are best treated as research-oriented descriptive measures.

Declarations

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest
The author declares no conflicts of interest.

Ethics and Consent
This secondary analysis used only de-identified data from a public repository and involved no participant contact. The repository did not document the originating ethics committee, approval number, or consent process.

Data and Code Availability
No new participant data were collected. This study performed a secondary analysis of the publicly available Q-CHAT dataset, version 2, available from Mendeley Data (doi:10.17632/tmpkt2mfkg.2). The Python analysis code and derived machine-readable outputs are available from the corresponding author on reasonable request.

Acknowledgments
None

References

[1] Allison C, Baron-Cohen S, Wheelwright S, Charman T, Richler J, Pasco G, et al. The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18-24 months of age: preliminary report. J Autism Dev Disord. 2008;38(8):1414-1425. doi:10.1007/s10803-007-0509-7.

[2] Petrocchi S, Levante A, Lecciso F. Systematic review of Level 1 and Level 2 screening tools for autism spectrum disorders in toddlers. Brain Sci. 2020;10(3):180. doi:10.3390/brainsci10030180.

[3] Allison C, Matthews FE, Ruta L, Pasco G, Soufer R, Brayne C, et al. Quantitative Checklist for Autism in Toddlers (Q-CHAT). A population screening study with follow-up: the case for multiple time-point screening for autism. BMJ Paediatr Open. 2021; 5(1): e000700. doi:10.1136/bmjpo-2020-000700.

[4] Magiati I, Goh DA, Lim SJ, Gan DZQ, Leong JCL, Allison C, et al. The psychometric properties of the Quantitative-Checklist for Autism in Toddlers (Q-CHAT) as a measure of autistic traits in a community sample of Singaporean infants and toddlers. Mol Autism. 2015; 6: 40. doi:10.1186/s13229-015-0032-1.

[5] Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev. 2016; 41: 71-90. doi:10.1016/j.dr.2016.06.004.

[6] Niedźwiecka A, Pisula E, Domasiewicz Z. Q-CHAT scores of Polish toddlers with autism spectrum disorders and typically developing controls. Mendeley Data. 2020; version 2. doi:10.17632/tmpkt2mfkg.2.

[7] Niedźwiecka A, Pisula E. Symptoms of autism spectrum disorders measured by the Qualitative Checklist for Autism in Toddlers in a large sample of Polish toddlers. Int J Environ Res Public Health. 2022;19(5):3072. doi:10.3390/ijerph19053072.

[8] Mujeeb Rahman KK, Monica Subashini M. A deep neural network-based model for screening autism spectrum disorder using the Quantitative Checklist for Autism in Toddlers (QCHAT). J Autism Dev Disord. 2022;52(6):2732-2746. doi:10.1007/s10803-021-05141-2.

[9] Allison C, Auyeung B, Baron-Cohen S. Toward brief "Red Flags" for autism screening: the Short Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in Toddlers in 1,000 cases and 3,000 controls [corrected]. J Am Acad Child Adolesc Psychiatry. 2012;51(2):202-212.e7. doi:10.1016/j.jaac.2011.11.003.

[10] Ruta L, Chiarotti F, Arduino GM, Apicella F, Leonardi E, Maggio R, et al. Validation of the Quantitative Checklist for Autism in Toddlers in an Italian clinical sample of young children with autism and other developmental disorders. Front Psychiatry. 2019; 10: 488. doi:10.3389/fpsyt.2019.00488.
[11] Stevanović D. Quantitative Checklist for Autism in Toddlers (Q-CHAT): a psychometric study with Serbian toddlers. Res Autism Spectr Disord. 2021; 83: 101760. doi:10.1016/j.rasd.2021.101760.
[12] Sturner R, Howard B, Bergmann P, Attar S, Stewart-Artz L, Bet K, et al. Autism screening at 18 months of age: a comparison of the Q-CHAT-10 and M-CHAT screeners. Mol Autism. 2022;13(1):2. doi:10.1186/s13229-021-00480-4.
[13] Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999;282(11):1061-1066. doi:10.1001/jama.282.11.1061.
[14] Goh DA, Gan D, Kung J, Baron-Cohen S, Allison C, Chen H, et al. Child, maternal and demographic factors influencing caregiver-reported autistic trait symptomatology in toddlers. J Autism Dev Disord. 2018;48(4):1325-1337. doi:10.1007/s10803-018-3471-7.
[15] Wong HS, Huertas-Ceballos A, Cowan FM, Modi N. Evaluation of early childhood social-communication difficulties in children born preterm using the Quantitative Checklist for Autism in Toddlers. J Pediatr. 2014;164(1):26-33.e1. doi:10.1016/j.jpeds.2013.07.013.
[16] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 2007;4(10):e296. doi:10.1371/journal.pmed.0040296.
[17] Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. J Educ Stat. 1981;6(2):107-128. doi:10.3102/10769986006002107.
[18] Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika. 2009;74(1):107-120. doi:10.1007/s11336-008-9101-0.
[19] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.