Journal of Culture and Values in Education

https://cultureandvalues.org

E-ISSN: 2590-342X

Volume: 9 Issue: 3 2026

pp. 222-252

Development and Validation of a Psychometric Scale for Genetically Influenced Behavioral Patterns in School-Aged Children: An Exploratory Factor Analysis

Bander N. Alotaibi[*a], Ahmed A. Alabbasi[a], Eid G. Abo Hamza[b,c*], & Alaa Eldin A. Ayoub[d,e]

* Corresponding author:
Email: eidhamza@edu.tanta.edu.eg
a. Department of Special Education, Arabian Gulf University, Manama, Kingdom of Bahrain
b. College of Arts, Humanities and Social Sciences, University of Sharjah, Sharjah, UAE
c. Faculty of Education, Tanta University, Tanta, Egypt
d. Hurghada University, Hurghada City, Egypt
e. Aswan University, Aswan City, Egypt

Article Info
Received: June 07, 2026
Accepted: August 25, 2026
Published: September 29, 2026

10.46303/jcve.2026.47

How to cite
Alotaibi, B. N., Alabbasi, A. A., Abo Hamza, E. G., & Ayoub, A. E. A. (2026). Development and Validation of a Psychometric Scale for Genetically Influenced Behavioral Patterns in School-Aged Children: An Exploratory Factor Analysis. *Journal of Culture and Values in Education*, *9*(3), 222-252. https://doi.org/10.46303/jcve.2026.47

Copyright license
This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

ABSTRACT
The purpose of this study was to develop and validate a teacher-rated behavioral scale to identify behaviorally stable patterns in school-aged children that are conceptually grounded in behavioral genetics theory. The scale does not measure genetic variation, and no genetic or neurobiological data were collected. Using an instrument comprising 30 items aligned with the principles of behavioral genetics theory, data were collected from 124 teachers of kindergarten, primary, middle, and high school pupils in Riyadh. Using principal axis factoring with promax rotation, exploratory factor analysis revealed a three-factor structure composed of 14 items assessing externalizing behavioral dysregulation, 9 items assessing attentional dysregulation, and 7 items assessing social withdrawal and anxiety that explained 69.55% of the variance. The internal consistency of the scale was strong (α = .959), whereas the sampling adequacy was strongly supported, with KMO= .92 and a significant Bartlett's test ($p < .001$). According to the results of the ANOVA, educational level significantly differed: compared with primary and middle school pupils, kindergarten pupils scored higher. According to the regression analysis, educational level was the only significant predictor (B = -0.43, 95% CI [-0.77, -0.10], p = .012), although the model explained only 6.4% of the variance. Horn parallel analysis supported the factor retention. No evidence of convergent or discriminant validity was obtained, and furthermore, no content validity was quantified. The results therefore provide initial evidence of the internal structure and reliability of the scale within a Saudi sample rather than establishing it as a validated instrument.

KEYWORDS
child behavior; psychometric scale; school-based assessment; attentional dysregulation; externalizing behavioral dysregulation; social withdrawal; behavioral genetics

INTRODUCTION

Children exhibit a wide range of behavioral patterns that vary substantially in their origins and responsiveness to intervention. Typically, behaviors are learned through environmental exposure and social reinforcement, and children respond according to well-standardized behavioral strategies. Others, however, persist across time and contexts with relative resistance to conventional classroom management and counseling-based approaches. This persistence has led behavioral experts and educational practitioners to seek alternative explanations, focusing beyond the environment, to consider heritable, neurobiological behavior in childhood (Plomin et al., 2016; Rutter, 2006). Understanding why certain children continue to exhibit disruptive, withdrawn, or dysregulated behavior despite consistent and competent intervention efforts remains among the most pressing questions in contemporary educational psychology.

The past three decades have produced enormous evidence that individual differences in children's behavior are significantly affected by genetic endowment. Estimates of heritability coefficients in the 40% to 80% range have been consistently obtained in twin and adoption studies for attention-deficit/hyperactivity disorder (ADHD), conduct disorder, anxiety, and social withdrawal (Abo Hamza et al., 2020; Faraone et al., 2021; Polderman et al., 2015; Supena & Kusmawati, 2022). These approximations indicate that a substantial proportion of interindividual variation in the expression of behavior, control of emotion and sensitivity to instruction is attributable to variation in these individuals' genetic makeup. Translating behavioral genetic research into school-based assessment practices has been astonishingly slow, notwithstanding this strong scientific consensus. Teachers and school staff are frequently presented with situations in which they are faced with students whose behavior is difficult to manage and who do not have the tools they need to assess if the behaviors they are observing are environmentally- or neurobiology-driven. The difficulty of choosing interventions that can be well adapted to the real source of the problem is complicated by such a gap.

The inability to make this distinction has grave practical implications. When teachers interpret a genetically anchored attentional or social behavioral pattern as a motivational or disciplinary problem, they are likely to apply standard behavioral modification or reward-based strategies. These strategies produce minimal change and eventually lead to professional frustration. This trend was observed empirically by Sciutto et al. (2000) in a sample of teachers of considerable size, and the authors concluded that significant percentages of teachers had systematic misconceptions about the causes of attentional and hyperactive behaviors. This leads them to misattribute neurobiologically rooted traits to environmental or volitional causes. Perold et al. (2010) extended these findings to a non-Western educational context. The study demonstrated that teacher misperceptions of behavioral origin were associated with lower confidence in the selection of appropriate instructional accommodations. Samarakkody et al. (2010) argued that reliance on subjective teacher observation without standardized and theory-grounded measurement tools is the proximate cause of this systematic misclassification.

The problem is not limited to the Gulf region. Internationally, the psychometric literature on child behavioral assessment has produced several validated instruments. They include the Child Behavior Checklist (Achenbach & Rescorla, 2001), the Behavior Assessment System for Children (Reynolds & Kamphaus, 2015), the Conners Rating Scales (Conners, 2008), and the Strengths and Difficulties Questionnaire (Goodman, 2001). These instruments have made a significant impact in the field because they offer standardized, factor-analytically validated measures of children's behavioral and emotional problems. They share two important drawbacks, however. First, they were developed mainly for the purpose of categorizing observable behavioral symptoms with respect to the diagnostic criteria of the major classification systems (American Psychiatric Association, 2013; World Health Organization, 2019) or with respect to normative criteria. Second, there is no clear theoretical background that separates genetically stable behavior from behavior that can be modified owing to the environment. Even the most validated scales of ADHD ratings do not formally distinguish between the neurobiologically based and reactive aspects of attention and hyperactivity behavior (Barkley, 2015). This restricts their applicability in the case of choosing etiologically suitable interventions.

The theories in which the current instrument has been developed are based on several foundational studies. Gottesman and Gould (2003) proposed the idea of behavioral endophenotypes, which are “quantifiable behavioral traits” that fall on the causal pathway between genetic variation and observable clinical phenomena. Their model laid the groundwork for the use of stable, trait-like behavioral patterns observed in a school context to be a reliable indicator of underlying genetic influences, even when there are no molecular genetic data. Drawing on this work, Moffitt et al. (2008) used longitudinal data to demonstrate that children with early-onset and persistent behavioral problems are biologically different from those who manifest limited and situationally bound problems during adolescence. This distinction has a direct effect on treatment choice. More recently, Faraone et al. (2021) extracted more than 200 evidence-based findings and reported that the behavioral patterns of attention, hyperactivity and impulsivity are highly heritable in humans. They also contended that they are neurobiological and therefore different in kind from behaviorally similar ones that are only environmentally conditioned.

In the psychometric domain, Costello and Osborne (2005) established best practices for using exploratory factor analysis in behavioral scale development. They demonstrated that EFA is the most scientifically defensible method for uncovering the latent dimensional structure of behavioral rating data. Fabrigar et al. (1999) provided comprehensive guidelines for oblique rotation strategies, which are appropriate when theoretical factors are expected to be correlated, as is the case with behavioral dimensions sharing a genetic substrate. Together, these methodological contributions created the technical foundation for the present study. Using longitudinal data, Moffitt et al. (2008) demonstrated that early-onset behavioral difficulties tend to remain relatively stable and developmentally continuous into the school

years, reinforcing the developmental validity of measuring theoretically stable, genetically informed behavioral patterns during this period.

Despite these theoretical and methodological improvements, several critical gaps remain in the literature. First, there is no validated psychometric instrument specifically developed for the measurement of the dimensional structure of genetically influenced behavioral traits in the framework of ordinary classroom observation by teachers. Existing instruments are aimed at clinical diagnoses or general behavioral dimensions instead of theoretically differentiated etiological dimensions. Second, the Arabic language psychometric literature through child behavioral assessment is significantly underdeveloped. Emam et al. (2016) reported important psychometric limitations in available behavioral rating tools for the Arabic language. Most adapted instruments lack complete factor-analytic validation, age-reference normative data, and theoretical frameworks based on developmental psychopathology models. Eapen and Ghubash (2004) made the same observation that a lack of culturally calibrated screening instruments has left practitioners in the Gulf region heavily dependent on clinical impressions instead of standardized measurements. In line with this issue, Nigg (2017) and Cortese and Coghill (2018) suggested that etiologically informed, theoretically precise measurement is a prerequisite for selecting appropriate behavioral interventions; therefore, instruments to differentiate between neurobiologically stable and environmentally conditioned behaviors within the context of routine school-based observation are needed.

The present study was designed to address the abovementioned interconnected gaps. A 30-item behavioral rating scale was created on the basis of behavioral genetics theory and structured around a theoretical three-dimensional construct of genetically anchored or neurobiologically stable behavior, latent attentional and motivational behavior, and acquired overt behavior disruption. The scale was administered to a sample of Saudi elementary school teachers, and they rated those students who showed atypical patterns of behavior. Exploratory factor analysis was then used to determine the empirical latent structure of the instrument. The scale's reliability was examined, and the distribution of behavioral scores across demographic groups was examined. The resulting instrument fills a specific and documented gap in the cross-section of behavioral genetics theory, psychometric scale development, and educational practice in the Arabic region.

Accordingly, the present study was guided by the following objectives and research questions, which were specified a priori before the method was implemented. The primary objective was to develop and provide initial psychometric validation for a teacher-rated behavioral screening scale grounded in behavioral genetics theory. The following research questions were addressed in this study:

1. What is the underlying factor structure of the developed scale, and does it reflect the theoretically expected dimensions of genetically influenced behavior among school-aged children?
2. Is the scale a reliable measure of the behavioral dimensions it identifies?

3. Do children's behavioral scores differ according to gender and educational level?

LITERATURE REVIEW

Behavioral Genetics and Child Development

Behavioral genetics is a branch of science that studies the relative contribution of genetic and environmental factors to the variation in the behavior and psychological traits of individuals (Plomin et al., 2016). Its foundational methodologies include twin studies, adoption studies, and molecular genetic studies, such as genome-wide association studies. Decades of research on twins have shown that essentially all of the measurable psychological traits are highly heritable (Polderman et al., 2015). In the context of child development, this evidence has direct implications for the way in which practitioners conceptualize and respond to behavioral difficulties in school settings.

The heritability of attention deficit hyperactivity disorder has been estimated to be approximately 74% through a meta-analytic synthesis of studies on twins (Faraone et al., 2021). Conduct disorder and oppositional defiance disorder have heritability coefficients ranging from 40% to 65% (Moffitt et al., 2008). Social anxiety and withdrawal behaviors are 30% to 50% heritable, with genetic factors interacting with early environmental experiences (Scaini et al., 2014). These numbers suggest that a large proportion of the behavioral variability that occurs in the classroom has a genetic basis that deserves special approaches to assessment. It is not a counsel of therapeutic nihilism to understand this genetic contribution; rather, it provides an impetus for the development of more precise assessment tools that can discriminate between stable, trait-like patterns of behavior and behavior conditioned by the situation.

These heritability estimates are derived from twin and family designs, which provide estimates of the proportion of variance in the population attributable to genetic relatedness. They do not constitute evidence that any specific behavioral rating scale reflects genetic influences in a specific child. This distinction is important in practice, as evidenced by recent molecular genetic studies. The largest genome-wide association study of ADHD to date identified 27 significant genome-wide loci from a meta-analysis of more than 38,000 cases and 186,000 controls. This makes a significant contribution to the polygenetic model of the disorder (Demontis et al., 2023). In contrast, a parallel genome-wide association meta-analysis of childhood and adolescent internalizing symptoms in more than 64,000 children did not identify any genome-wide significant loci and estimated the SNP-based heritability to be just 1.66%, which is significantly lower than the 30–50% heritability reported by twin studies for this broader phenotype (Jami et al., 2022). This discrepancy is between the twin-based heritability and the actual amount of genetic signal that can be recovered from molecular data and has been termed the missing heritability problem. This is a current field of study in behavioral genetics and highlights that there is no simple or certain relation between the heritability of a trait at the population level and its detectability at a level of individual measured behaviors and certainly not at the level of a single child's score on a teacher-rated instrument.

A closely related set of fallacies warns against using heritability as a criterion for defining a trait or for determining its cause in an individual. Turkheimer (2000) codified this worry into the third law of behavior genetics. A significant portion of the variance in complex human behavioral characteristics can be explained neither by genes nor by the shared family environment but rather by nonshared environmental experiences and by measurement error not broken into a causal explanation by the heritability statistics for any individual. Moore and Shenk (2017) further report that heritability does not imply the degree of genetic determinism. It does not explain the causes of this, and it does not divide the relative effects of genes and the environment on any one person's behavior. None of these arguments exclude the possibility of heritability of behavioral traits at the population level but are warning against the danger of an inferential step from population heritability to the individual level of measurement by genes. Because of this, it is directly relevant to the interpretation of the results of the current instrument and is discussed again in the Discussion section.

Gottesman and Gould (2003) proposed the concept of endophenotypes. It is defined as quantifiable biological or behavioral traits found along the route from genetic to observable clinical phenomena. This concept has been very influential in the construction of dimensional behavioral assessment tools. In the context of school children, endophenotypical markers for genetically influenced traits include impaired executive function, attentional dysregulation, motor hyperactivity and social information processing difficulties. Rating scales that reflect these dimensions are a useful link between genetic science and educational practice (Barkley, 2015). The developmental continuity of such traits into the school years is supported by longitudinal evidence that early-onset behavioral difficulties are relatively stable over time (Moffitt et al., 2008), reinforcing the theoretical basis for the present instrument.

Classification of Behavioral Patterns in Children

Theoretical organization of human behavior in childhood has been made into several classifications. The tripartite model, which distinguishes between genetic or neurobiologically anchored behavior, latent behavior and acquired overt behavior, is a prominent model in behavior science (Beyazıt et al., 2023; Elsantil & Abo Hamza, 2021; Mash & Wolfe, 2019; Plomin et al., 2016). Genetic or neurobiologically anchored behavior manifests early in development. This is shown to be cross-situational consistency and relative resistance to conventional behavioral modification. Some examples are chronic motor hyperactivity, stereotyped or repetitive movements, social withdrawal from an early age, and emotional dysregulation without obvious environmental causes. These traits are characterized as being temporally stable and partially independent of reinforcement contingency (Barkley, 2015; Rutter, 2006). Behavioral genetics research has been quite supportive of the presence of heritable behavioral phenotypes of this type, and they have consistently been found using the biometric modeling of twins.

Latent behavior comprises behavioral tendencies that are not continuously observable but are manifested in certain situational contexts. These can be covert anxiety, inwardly

directed negative effects, and attentional difficulties that are manifested in mainly cognitively demanding tasks. Latent behavioral traits may reflect underlying motivational, affective or executive function deficits that are not readily evident to casual observation but become clear under conditions of structured assessment (Gottesman & Gould, 2003).

Acquiring overt behavior involves learned behavioral patterns that are influenced by environmental reinforcement, modeling and social learning processes. Following Bandura's (1977) social learning theory, these behaviors are more plastic and responsive to behavioral interventions. They include classroom disruption that is inadvertently reinforced by teacher attention, imitation of aggressive peer behavior and rule-breaking behavior that is associated with particular social contexts. The difference between these three types of behavior has important consequences for the design of interventions. Practitioners who confuse neurobiologically anchored traits with acquired behaviors risk implementing ineffective or counterproductive interventions (Mash & Wolfe, 2019).

Psychometric Assessment of Child Behavior

Over the past fifty years, there has been a significant yield of rating scales related to child behavioral assessment. These include the Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2001; Seleem et al., 2023), the Behavior Assessment System for Children (BASC-3) (Reynolds & Kamphaus, 2015), and the Conners Rating Scales (Conners, 2008). They are instruments that measure a wide variety of behavioral and emotional problems in a standardized, normed and factor-analytic manner. The Strengths and Difficulties Questionnaire (SDQ) (Goodman, 2001) provides a short, 5-factor alternative that has been widely adapted for use internationally, including in Arabic.

Cross-cultural and regional variations in behavioral presentation are well documented and have important implications for scale development (Achenbach et al., 2008). Studies conducted in cross-national and Arabic-region contexts have documented culturally specific patterns of behavioral expression and help-seeking that differ meaningfully from Western norms (Achenbach et al., 2008; Eapen & Ghubash, 2004; Seleem et al., 2023). Eapen and Ghubash (2004) reported that caregivers in the United Arab Emirates relied predominantly on informal support networks rather than formal clinical services when they responded to children’s behavioral difficulties. These findings underscore the need for locally developed and validated assessment instruments rather than direct translations of Western tools, which may not adequately capture the culturally specific behavioral presentation norms of Gulf-region student populations. The limitations of the existing Western instruments in this region are not just about language. Emam et al. (2016) evaluated the Arabic Strengths and Difficulties Questionnaire. They indicated that adapted instruments are often not completely revalidated on Arabic samples in terms of factor analysis. In addition, region-specific normative benchmarks and item content, calibrated to local classroom and family expectations, are necessary because factor structures developed on Western normative data are not necessarily recovered. Eapen and Ghubash (2004) also demonstrated that the context of help-seeking and reporting in the

Gulf is not as assumed in the west. This most limits the interpretation of directly transposed instruments such as CBCL, BASC-3, Conners and SDQ. Not only a lack of an etiological framework in general but also these specific psychometric and contextual gaps. This was a direct impetus for the present decision to develop and factor-analytically validate a new theory-based instrument on a local sample instead of adapting an instrument based on a Western sample.

A recurring theme across the assessment literature is that accurate, etiologically informed characterization of a child's difficulties must precede intervention selection. Nigg (2017) argued that knowing the precise nature of a child's regulatory difficulty is essential before any intervention can be meaningfully designed. Zelazo and Carlson (2020) similarly reported that behavioral interventions are most effective when the underlying neurobiological deficit is accurately identified beforehand. This conclusion is supported by the work of Cortese and Coghill (2018), who stated that etiologically informed assessment is the basis of effective behavioral support. From all three perspectives, one thing becomes clear: psychometrically sound behavioral assessment instruments are not a methodological luxury but are a scientific and practical necessity.

Factor Analysis in Behavioral Scale Development

Factor analysis has been widely employed in the development of behavioral and psychological instruments of measurement. EFA reveals the number of latent constructs underlying a set of items that are observed. EFA is useful in the identification of item clusters that fit together as indicators of a common factor. It also helps to eliminate the items that do not contribute meaningfully to any factor (Tabachnick & Fidell, 2019). The resulting structure of the factors is the empirical basis for subscale construction, scoring and theoretical interpretation.

Costello and Osborne (2005) identified EFA as the most appropriate method for uncovering the latent structure of behavioral constructs; however, one must emphasize the importance of adequate sample size, appropriate extraction criteria, and rotation strategy. Fabrigar et al. (1999) and Preacher and MacCallum (2003) subsequently developed best-practice guidelines for EFA in scale development research. These guidelines include the use of oblique rotation methods when factors are theoretically expected to correlate. They further recommend that factor retention be guided by the convergent use of several criteria, such as the Kaiser eigenvalue-greater-than-one rule, visual inspection of the scree plot for the point of inflection, and the cumulative proportion of variance explained. These are major criteria for interpreting alongside the theoretical and substantive interpretability of the resulting solution, with factor loadings above 0.35 treated as indicative of meaningful item–factor relationships. Tabachnick and Fidell (2019) reported that KMO values above 0.80 indicate meritorious sampling adequacy and that Bartlett's test should be significant at $p < .05$.

In the tradition of behavioral genetics, factor analytic approaches have been used to identify behavioral phenotypes that correspond to heritable endophenotypes (Gottesman & Gould, 2003). The heritability of a behavioral factor identified through EFA can be estimated by applying the twin methodology to factor scores that link the psychometric and genetic levels of

analysis. This approach has been applied successfully in studies of executive function (Miyake et al., 2000), emotional regulation (Scaini et al., 2014), and attentional control (Faraone et al., 2021).

RESEARCH METHOD

Research Design

A descriptive-analytical cross-sectional research design was used in this study. This design is suitable for scale development and initial psychometric validation. It enables the study of the structure of factors and internal consistency of the instrument in a single data collection period (Kline, 2016). The study followed the major phases of the scale development process as outlined by DeVellis (2016). The item generation is based on a theoretical review followed by the review of an expert panel and statistical refinement of the items through EFA. This sequence ensures that the resulting instrument has both theoretical coherence and empirical support.

Participants and Sampling

The targeted population consisted of male and female teachers who teach at all levels (kindergarten, primary, middle and high) in Riyadh city, Saudi Arabia. The target population will therefore be in terms of school-aged children at these four levels, not just the elementary schools themselves. Throughout, school-aged children are used in this way. Teachers were in direct interaction with the students who exhibited atypical behavior patterns. Each teacher completed the behavioral rating scale for one focal student who he/she felt had behavior problems of note. The demographic data of the sample are summarized in Table 1.

The sample includes students enrolled in both general education and special education programs (n = 70 and n = 54, respectively). Enrollment in special education programming in this setting extends the practical age range of “school-aged”. Every case above 18 years in the full sample (n = 9, ages 18–41) was enrolled in a special education program.

Nonprobability convenience sampling was used for the recruitment of teachers. The schools in Riyadh were approached and were accessible to the research team. Those teachers in those schools who were available and willing to participate were represented. No random sampling of schools or teachers within schools was performed. The focal students were not randomly selected but were purposively selected as those who were identified by each teacher as having some noticeable behavioral difficulty within each participating classroom. The sample is therefore not epidemiologically representative of the student population in Riyadh, and the reported scores can be interpreted only as describing a behaviorally referred sample and not as a general population norm.

Participation was voluntary. The teacher was informed of the goals of the study, advised that he/she did not have to participate in the study and could withdraw at any time without providing a reason and without any penalty, and was asked to sign a written informed consent before completing the rating scale. There were 124 rating forms returned. The 124 complete cases (N = 124) were used for all the analyses below.

Table 1.

Demographic Characteristics of the Sample (N = 124)

Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	81	65.3
	Female	43	34.7
Educational Level	Kindergarten	7	5.6
	Primary School	90	72.6
	Middle School	15	12.1
	High School	12	9.7
Educational Program	General Education	70	56.5
	Special Education	54	43.5
Age (Years)	M = 11.44, SD = 5.54, Range: 5 to 41		

Study instrument

The behavioral rating scale was designed specifically for this study. The construct it measures is defined as teacher-observed behavioral patterns that are theorized, on the basis of behavioral genetics research, to be relatively stable, early-emerging, and comparatively resistant to routine classroom intervention. The instrument does not measure genetic variation. No genotyping, biomarker, family history, or other neurobiological data were collected, and no inference about the genetic etiology of any individual child behavior could be drawn from its scores. Wording such as genetically influenced behavior is therefore used throughout as shorthand for behaviorally stable traits that are conceptually grounded in behavioral genetics theory and not as a claim that genetic variation itself was assessed. This scale is based on a systematic review of a range of theoretical information from behavioral genetics, neuropsychology, applied behavior analysis, and empirical studies of child behavioral assessment (Barkley, 2015; Mash & Wolfe, 2019; Plomin et al., 2016; Rutter, 2006). The item pool consisted of 30 behavioral descriptors grouped into three theoretical subscales:

Items 1 to 10: Addressing behavior with genetic or neurobiologically anchored (e.g., hyperactivity, stereotyped movements, early social withdrawal).

Items 11 to 20: Targeting latent attentional and motivational behavior (e.g., attention difficulties, hesitation, lack of task engagement).

Items 21 to 30: Targeting acquired overt behavioral disruption (e.g., imitation of inappropriate peer behavior, refusal of instructions, classroom disruption).

Items were written to be directly observable by classroom teachers and reviewed by a panel of 5 experts in the field of educational psychology and special education prior to data collection. The experts were asked to comment qualitatively on the clarity, observability, and theoretical alignment of each item. Their feedback was used to reword ambiguous items and to improve ecological validity. Responses were recorded on a five-point Likert scale (1 = never; 2 =

rarely; 3 = sometimes; 4 = often; 5 = always). This format makes it possible to quantitatively grade such that it can be factor analyzed and reliability estimated (DeVellis, 2016).

Data collection

Academic data were collected for the years 2025 to 2026. The participating teachers were given a standardized briefing on the study and rating procedure. Written informed consent was obtained from all participants prior to data collection. Teachers were asked to complete the scale for one focal student in their class who was recognized by them as having atypical behavioral patterns. The rating task was to be completed individually within the school setting and returned to the research team within one week of distribution.

Data Analysis

All the statistical analyses were performed with the use of IBM SPSS Statistics (Version 20). The analytic strategy was organized into four phases and seven sequential stages to ensure rigorous psychometric assessment and hypothesis testing.

Phase I: Psychometric and Distributional Evaluation

Stage 1: Descriptive Statistics and Data Screening

The initial stage involved a comprehensive evaluation of the distributional properties of all 30 measurement items. Means (M), standard deviations (SD), and skewness and kurtosis coefficients were calculated for each item. This stage was necessary for two reasons: to describe the central tendency and the spread of item responses and to determine whether there are any dangers of nonnormality prior to factor analysis. Following the criteria recommended by Kline (2016), absolute values of skewness greater than 2.0 or kurtosis greater than 7.0 were considered to represent extreme nonnormal values and are presented for further inspection.

Stage 2: Assessment of Sampling Adequacy and Factorability

Before factor extraction was performed, the adequacy of the correlation matrix for factor extraction was assessed through two complementary indices. The Kaiser–Meyer–Olkin (KMO) index of sampling adequacy was interpreted in line with the criteria developed by Tabachnick and Fidell (2019), where .60 and higher is acceptable and values above .80 are meritorious. Another test of sphericity was also applied: Bartlett's test. A statistically significant finding ($p < .05$) was necessary to reject the null hypothesis that the population correlation matrix is an identity matrix, which is necessary to confirm that there are sufficient correlations between items to warrant factor analysis.

Phase II: Factor Structure and Dimensionality

Stage 3: Exploratory Factor Analysis

The latent dimensional structure of the behavioral scale was determined by principal axis factoring (PAF). Promax oblique rotation was used on the theoretical assumption that the underlying behavioral dimensions would be correlated because of their shared genetic and neurobiological substrate. Four convergent criteria were used to determine the number of factors to retain. The first was the Kaiser criterion, which is the factor with eigenvalues > 1.0. The second was a visual examination of the scree plot; the point of inflection was used as an

indicator of retention. The third was the accumulation of the variance accounted for by the solution extracted. Another criterion was added: Horn parallel analysis. The observed correlation matrix eigenvalues were compared with the eigenvalues of 1000 random data matrices of the same size (124 cases, 30 variables) following the procedure suggested by Fabrigar et al. (1999) and Costello and Osborne (2005). If the eigenvalue of a factor obtained from the sample data was greater than the eigenvalue of the same factor obtained from the random data, then it was deemed supported. The distributions of the random eigenvalues were studied both at the mean and at the 95th percentile. Since these two decision rules can differ for the marginal factors, the more conservative 95th percentile rule was kept as a sensitivity check. These statistical criteria were weighed, along with the theoretical and substantive interpretability of the resulting solution, in deciding on dimensionality. Factor loadings of .35 or higher were interpreted as meaningful indicators of item–factor relationships as per the guidelines of Fabrigar et al. (1999).

Phase III: Reliability and Item-Level Analysis

Stage 4: Internal Consistency/Reliability

The reliability of each extracted factor subscale and the full 30-item scale was estimated using Cronbach's alpha (α). To ensure the precision of the reliability estimates and to facilitate comparison with established benchmarks, 95% confidence intervals were computed for all alpha coefficients using the Feldt method (Feldt et al., 1987). Alpha values of .70 and above were considered acceptable for research use, with values of .80 and above considered adequate for screening purposes (Kline, 2016; Nunnally, 1978).

Stage 5: Item discrimination and scale refinement

To ensure that the items on each subscale added value, item-level diagnostics were performed. The following two statistics were assessed for each item. Corrected item–total correlations (CITCs) were obtained by correlating each item with the total score of the scale excluding that item. Values less than .30 were interpreted as indicators of poor item discrimination (Kline, 2016). Alpha-if-item-deleted values were also examined to determine whether deletion of the item would result in a significant increase in overall scale reliability. Low CITCs and high alpha-if-item-deleted values were marked for consideration of item deletion.

Phase IV: Relational and Group-based Analyses

Stage 6: Interfactor Correlations and Subscale Characterization

Bivariate Pearson's correlations of subscale mean scores were used to determine the relationships among the three dimensions of behavior identified. Means, standard deviations and score ranges are reported for each subscale. This stage was intended to describe the amount of overlap and the unique aspects of the three behavioral constructs, and the interfactor correlations offered empirical support for the oblique rotation (as opposed to orthogonal rotation) that was used in Stage 3.

Stage 7: Demographic Comparisons and Inferential Testing

In the final stage, we examined how the behavioral scores varied across the demographic subgroups. Gender differences in total behavioral scores were examined using an independent samples t test, with Cohen's d reported as a measure of effect size. Educational level differences were evaluated with one-way analysis of variance (ANOVA) with pairwise post hoc testing with Bonferroni adjustments to control for Type I error inflation (familywise) for the six possible group combinations. Finally, multiple linear regression was used to estimate the joint predictive utility of demographic variables, that is, age, sex and educational level, for total behavioral scores. Unstandardized (B) and standardized (b) regression coefficients, standard errors, t statistics, and corresponding P values are presented along with the overall model fit statistics (R2, adjusted R2, and model F-ratio).

Ethical Considerations

The study was conducted in compliance with institutional and national ethical standards and the Declaration of Helsinki (World Medical Association, 2013). Written informed consent was obtained from all the teacher participants. Student confidentiality was preserved throughout by using numbers instead of names. The research protocol was reviewed and approved following the institutional policies of Arabian Gulf University.

RESULTS

Descriptive Statistics

Table 2 shows the descriptive statistics for all 30 items of the behavioral scale. The mean range of values of the items was 1.96–2.69. This suggests generally moderate levels of behavioral endorsement frequencies across the sample. The standard deviations ranged from 1.27 to 1.55, indicating a good distributional spread for factor analysis. Skewness ranged from 0.22 to 1.21, and kurtosis ranged from -1.34 to 0.36, suggesting that there was a slight positive skew for most items. This is in line with the overall lower behavioral severity in school populations. No item had extreme distributional deviations that would preclude the use of factor analysis or reliability estimation.

Sampling Adequacy and Factorability

The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was computed as 0.92. This value exceeds the recommended threshold of 0.80 and is classified as excellent (Tabachnick & Fidell, 2019). Bartlett's test of sphericity was statistically significant (chi-square = 3511.45, df = 435, $p < .001$), confirming that the interitem correlation matrix differed significantly from an identity matrix and was therefore appropriate for factor analysis. These indicators provided strong justification for proceeding with EFA.

Exploratory Factor Analysis (EFA)

Principal axis factoring with promax oblique rotation was conducted. The Kaiser eigenvalue-greater-than-one criterion indicated three factors with eigenvalues of 14.01, 5.03, and 1.82. Together, these three factors accounted for 69.55% of the total variance, with the first factor

alone explaining 46.71%. The scree plot was consistent with a three-factor interpretation, showing a prominent elbow after the first factor and a secondary inflection between factors three and four. Horn parallel analysis was subsequently performed on 1,000 matrices of random data with the same number of cases and variables (124×30). Both decision rules yielded observed eigenvalues (14.01 and 5.03) that far surpassed those of the random data counterparts and that were clearly supported. The third factor also barely fulfilled the mean criterion, with an observed eigenvalue equal to 1.82, compared with the mean for a random matrix of 1.77. However, it was not supported under the more conservative rule, as the third factor was just below the 95% random eigenvalue of 1.87. The fourth factor was strongly rejected under both rules, with observed values of 0.98, compared with random values of 1.67 and 1.75. Therefore, parallel analysis indicates that two factors can be accepted with confidence and that the third factor is located at the decision boundary, which is a more conservative result than that of the Kaiser criterion alone. Eigenvalues and variance statistics for the retained factors are given in Table 3, while the complete promax rotation pattern matrix is given in Table 4.

Table 3.

Eigenvalues and Variance Explained for the Three Retained Factors

Factor	Label	Eigenvalue	% Variance	Cumulative %
1	Externalizing Behavioral Dysregulation	14.01	46.71	46.71
2	Attentional Dysregulation	5.03	16.77	63.48
3	Social Withdrawal and Anxiety	1.82	6.07	69.55

A comprehensive review of the entire pattern matrix reveals that the extent of cross-loading is greater than what a two-item description implies at the |0.35| threshold. Eight (08) of the 30 items (26.7%) load at or above |0.35| on two or more factors: “Avoids social interaction” (Item 2; F2 = 0.462, F3 = 0.494), “Anxiety or social withdrawal” (Item 3; F2 = 0.383, F3 = 0.710), “Persistent isolation” (Item 4; F2 = 0.406, F3 = 0.708), “Avoids eye contact” (Item 7; F2 = 0.488, F3 = 0.639), “Unpredictable behavior” (Item 9; F1 = 0.701, F3 = 0.356), “Internal tension” (Item 12; F2 = 0.473, F3 = 0.501), “Mental distraction in class” (Item 18; F1 = 0.410, F2 = 0.654), and “Withdraws from learning situations” (Item 26; F1 = 0.560, F2 = 0.541). This represents a materially greater rate of cross-loading than the impression created by referring to Items 2 and 26 alone. This is directly pertinent to the discriminant validity of the three-factor solution because approximately 25% of the items in the item pool do not present a pure factor pattern. Six of the eight cross-loading items (3, 4, 7, 9, 12, and 18) are related to Factor 3, which is consistent with the significant correlations of Factor 3 with Factor 1 (r = 0.420) and Factor 2 (r = 0.691) presented in the Interfactor Correlations section.

The internalizing dimension is the dimension that is the least clearly separated of the three dimensions. However, all eight were kept on one factor for purposes of scoring, and one decision rule was used for all eight. The items were allocated to the factor in which they

obtained their highest numerical loading, keeping in mind their content and the theoretical definition of the factor. In addition, if an item had a primary loading on a factor, it was not scored on a factor on which it did not have its primary loading. According to this rule, “Withdraws from learning situations” (Item 26) was very close to being assigned to Factor 1 rather than Factor 2 (0.019), and “Avoids social interaction” (Item 2) was very close to being assigned to Factor 3 rather than Factor 2 (0.032). Content considerations reinforced the numerical decision in both cases because the avoidance of social interaction and disruptive withdrawal both conceptually fit with the internalizing and disruptive-withdrawal constructs on which they were retained, respectively. The other six cross-loading items exhibited greater differences in their primary and secondary loadings. They were less ambiguously assigned, but they still exhibited meaningful cross-loading and were not just structured. In summary, this pattern suggests that the three-factor solution does not achieve complete item-level discriminant separation, especially with respect to Factor 3. It is not considered a labeling detail but a substantive limitation in the current version of the scale and is discussed in more detail later in the Discussion and Limitations.

Interfactor Correlations

Because promax oblique rotation was used, interfactor correlations were examined to assess the degree of convergence among the three behavioral dimensions. These correlations are provided in Table 5. All three factors were significantly positively correlated with each other ($p < .001$). This justifies the use of oblique rotation instead of orthogonal rotation. The strongest correlation was observed between Attentional Dysregulation and Social Withdrawal and Anxiety (r=.691). This implies that children who suffer from severe attention problems are also likely to be socially withdrawn. The correlations between externalizing behavior dysregulation and the other two factors were moderate (r = 0.491 and r = 0.420, respectively). These findings indicate that although the three dimensions share variance, they nevertheless represent meaningfully distinct behavioral constructs.

Table 5.

Interfactor Correlations among the Three Behavioral Subscales (N = 124)

Subscale	F1: Beh. Dysreg.	F2: Atten. Dysreg.	F3: Soc. W/A
F1: Externalizing Behavioral Dysregulation	—	0.491***	0.420***
F2: Attentional Dysregulation	0.491***	—	0.691***
F3: Social Withdrawal and Anxiety	0.420***	0.691***	—

Note. All correlations are Pearson product–moment correlations based on subscale mean scores. *** $p < .001$ (two-tailed). F1 = Externalizing Behavioral Dysregulation; F2 = Attentional Dysregulation; F3 = Social Withdrawal and Anxiety.

Reliability Analysis

Internal consistency reliability was calculated with Cronbach's alpha with 95% confidence intervals calculated using the Feldt method (Feldt et al., 1987). The reliability statistics and subscale descriptive information are presented in Table 6. All reliability coefficients greatly exceeded the conventional threshold of 0.70 that has been recommended for research instruments (Nunnally, 1978), and all exceeded the 0.80 threshold recommended for clinical screening instruments (Kline, 2016).

Table 6.

Subscale and Full-Scale Reliability Statistics with 95% Confidence Intervals

Subscale	Items	Mean	SD	Min	Max	alpha [95% CI]
F1: Externalizing Behavioral Dysregulation	14	2.395	1.192	1.00	5.00	0.962 [0.952, 0.971]
F2: Attentional Dysregulation	9	2.561	1.171	1.00	5.00	0.949 [0.935, 0.962]
F3: Social Withdrawal and Anxiety	7	2.185	1.032	1.00	5.00	0.882 [0.847, 0.911]
Full Scale (30 items)	**30**	**2.416**	**0.952**	**1.00**	**5.00**	**0.959 [0.948, 0.968]**

Note. CI = confidence interval estimated using the Feldt method. The mean and SD are calculated from the subscale item means. The full-scale row is based on all 30 items.

Corrected Item–Total Correlations

Corrected item–total correlations (CITCs) were calculated for each of the 30 items to the sum score for the 30 items. All the items had CITCs with a mean of > .30 (the recommended minimum CITC), ranging from 0.402 (Item 2: Avoid social interaction) to 0.786 (Item 26: Withdraws from learning situations). None of the alpha-if-item-deleted values were larger than the overall scale alpha, indicating that all the items contributed positively to the homogeneity of the scale. The complete results in terms of CITC and alpha-if-item deletion are provided in Table 7.

Demographic Group Comparisons

4.7.1 Gender Differences

Independent-samples t tests were used to compare total behavioral scores between male and female students. Male students (M = 2.476, SD = 0.975, n = 81) scored slightly higher than female students did (M = 2.244, SD = 0.909, n = 43). This difference did not reach statistical significance, t(122) = 1.29, p = .199. The mean difference was 0.232 scale points, 95% CI [-0.124, 0.588]; because this interval includes zero, the direction of any sex difference is not determined by these data. The effect size was small (Cohen's d = 0.24), indicating limited practical significance. These findings suggest that the behavioral patterns captured by the scale are not strongly differentiated by student gender in this sample.

4.7.2 Educational Level Differences and Post Hoc Analysis

One-way analysis of variance (ANOVA) was used to analyze differences in total behavior scores by educational level (four levels). The result was statistically significant, F(3, 120) = 7.95, p < .001. The means, standard deviations, 95% confidence intervals, and pairwise comparisons based on the Bonferroni correction are listed in Table 8. Kindergarten students had the highest mean score (M = 3.781, SD = 1.272, 95% CI [2.605, 4.957]), followed by high school (M = 2.683, SD = 1.020, 95% CI [2.035, 3.332]), primary school (M = 2.334, SD = 0.857, 95% CI [2.155, 2.514]), and middle school students (M = 1.889, SD = 0.711, 95% CI [1.495, 2.283]). The confidence interval for the kindergarten mean is much larger than that for all the others because of the small number of cases in that cell (n = 7). This inaccuracy should be kept in mind in any contrast made with the kindergarten group.

Table 8.

Differences in Behavioral Scores by Educational Level with Post Hoc Comparisons

Educational Level	n	M	SD	Post-Hoc Comparisons (Bonferroni)
Kindergarten	7	**3.781**	1.272	Sig. higher than Primary*** (d = 1.63) and Middle School*** (d = 2.07)
Primary School	90	2.334	0.857	Not sig. different from Middle School or High School
Middle School	15	1.889	0.711	Lowest mean; sig. lower than Kindergarten***
High School	12	2.683	1.020	Not sig. different from Primary or Middle School

ANOVA: F(3, 120) = 7.95, p < .001

Note. Pairwise comparisons were performed using the Bonferroni correction across six pairs. *** p < .001. Effect sizes are Cohen's d based on pooled variance. Significance is reported after the Bonferroni correction.

Multiple Regression Analysis

Multiple linear regression was used to test the joint effects of age, sex and educational level on total behavioral scores. The overall model was significant, F(3,120) = 2.74, p = .046, R^2 = 0.064, adjusted R^2 = 0.041. The model accounted for 6.4% of the variance in the behavioral scores. The regression coefficients are available in Table 9. Educational level was the only statistically significant predictor (B = -0.431, SE = 0.169, beta = -0.320, t = -2.55, p = .012, 95% CI [-0.766, -0.096]). This suggests that behavioral scores decreased as school levels increased. The confidence intervals for age (B = 0.038, 95% CI [-0.004, 0.079], p = .078) and sex (B = 0.329, 95% CI [-0.034, 0.691], p = .075) both include zero. Thus, neither predictor was significant, and the direction of their effects was not determined by these data. Notably, the variance accounted for by the model is only 6.4%, and the confidence interval of the educational level is wide; therefore, this coefficient is considered a preliminary estimate and not an exact one.

Table 9.

Multiple Regression Coefficients Predict Total Behavioral Scores from Demographic Variables

Predictor	B	SE	beta	t	p	95% CI
Intercept	2.725	0.286	—	9.534	< .001	[2.159, 3.290]
Age	0.038	0.021	0.218	1.776	.078	[-0.004, 0.079]
Gender (Male = 1)	0.329	0.183	0.164	1.794	.075	[-0.034, 0.691]
Educational Level	**-0.431**	0.169	-0.320	-2.550	**.012***	**[-0.766, -0.096]**

Note. B = unstandardized regression coefficient; SE = standard error; beta = standardized coefficient; 95% CI = 95% confidence interval for B. R squared = .064; Adjusted R squared = .041; F(3, 120) = 2.74, p = .046. Educational level coded: 1 = Kindergarten, 2 = Primary, 3 = Middle School, 4 = High School. * p < .05.

DISCUSSION

In this study, a psychometric scale for the assessment of genetic-influenced behavioral patterns among school-aged children was developed and validated using EFA. The results strongly support an empirical three-factor structural model, i.e., externalizing behavioral dysregulation, attention dysregulation, and social withdrawal and anxiety. The first factor, externalizing behavioral dysregulation, described a group of externalizing behaviors that included repetitive and stereotyped behaviors, the deliberate provocation of peers, classroom disruption, noncompliance, and uncontrollable motor behavior. This pattern is consistent with the externalizing spectrum described by Achenbach and Rescorla (2001) in the Child Behavior Checklist. This is mapped onto behavioral phenotypes that are associated with conduct problems and hyperactive-impulsive presentations of ADHD, which are shown to have moderate-to-high heritability of approximately 65% to 80% (Faraone et al., 2021; Moffitt et al., 2008). The high factor loadings on items related to deliberate provocation, repetitive behaviors, and unauthorized movement suggest that this factor represents a particularly overt, action-oriented dimension of genetically influenced behavior.

The second factor, attention dysregulation, was a group of inattentive behaviors, such as difficulty concentrating, hesitation before starting a task, mental distraction, slow cognitive processing and reduced participation. This profile corresponds closely to the inattentive dimension of ADHD and to executive function deficits identified as core heritable endophenotypes in behavioral genetics research (Barkley, 2015; Miyake et al., 2000). The very high factor loadings on items related to hesitation, slow responses, and difficulty concentrating. All the items exceeding a factor loading of 0.80 indicate that these items capture a particularly coherent and discriminable construct within the broader behavioral landscape of classroom behavior.

The third factor, social withdrawal and anxiety, included items reflecting persistent social avoidance, early-onset anxiety, avoidance of eye contact, sensory hypersensitivity, and negative self-evaluations. This dimension is consistent with the internalization of behavioral profiles related to behavioral inhibition temperament and social anxiety disorder. The results revealed substantial heritability in twin studies (Anttila et al., 2018; Scaini et al., 2014). The fact that this factor has emerged as a distinct dimension provides support for the conceptual validity of separating the internalizing and externalizing components of a scale that targets genetically influenced behavior. Importantly, there was a strong correlation between Factor 2 and Factor 3 (r = 0.691), which indicates that attentional problems and social withdrawal often cooccur in this population. These findings are consistent with patterns of comorbidity noted in the clinical literature concerning the coexistence of both AD/HD and anxiety disorders (Mash & Wolfe, 2019).

The reliability coefficients generated in this study show favorable comparisons with data from similar behavioral assessment instruments generated in other countries. The overall Cronbach's alpha is 0.959, which is higher than the alpha of 0.91. This is usually reported in terms of the CBCL externalizing scale (Achenbach & Rescorla, 2001) and is comparable to the reliability estimates for the BASC-3 subscales, which range from 0.80 to 0.97 (Reynolds & Kamphaus, 2015). The subscale alphas for Externalizing Behavioral Dysregulation (0.962) and Attentional Dysregulation (0.949) are higher than the alphas for the Conners' 3 teacher rating scales (Conners, 2008), providing evidence that the newly developed scale is at least as good as the established Western instruments in terms of internal consistency.

However, these reliability data should be viewed in light of a limitation of the present study. No convergent or discriminant validity evidence was collected in the present study. There was no correlation between scale scores and any existing measures of the same constructs, including the Conners scales, the CBCL, the BASC-3, the SDQ and the Arabic ADHD Rating Scale-5 normed for Saudi Arabia by Pierce et al. (2025). Additionally, these were not evaluated with respect to measures of unrelated constructs to establish discriminant separation. In this sense, the construct validity of the instrument is currently based only on its internal structure as well as the theoretical coherence of the content of the items.

Exploratory factor analysis at the development phase revealed a reasonable and plausible internal structure with acceptable reliability. It does not claim that the scale is a “true” measure, in the sense that it is usually understood. The very high alpha values presented here on their own should not be interpreted as a sign of confidence but rather caution: an alpha of 0.959 for 30 items. When combined with interfactor correlations as high as r = 0.691, it is also highly indicative of high item redundancy and a strong general factor across the three dimensions. Therefore, the scale is best considered to be structurally and reliably validated in a Saudi sample and not as a validated instrument for the assessment of genetically influenced behavior. Its subscale scores are tentative, with the meaning of these scores being awaited

convergent or discriminant evidence; this qualification extends to the practical applications that follow below.

The first factor, externalizing behavioral dysregulation, accounts for 46.71% of the variance, with an eigenvalue of 14.01, which is consistent with the findings of factor-analytical studies of child behavioral rating scales. In this context, the domain of behavioral difficulty rated by teachers is usually the most variable dimension, namely, externalizing (Achenbach & Rescorla, 2001). According to an international consensus review (Faraone et al., 2021), hyperactive–impulsive and disruptive behavioral presentations are among the most consistently identified and most heritable dimensions of child behavior. This factor should be noted; however, for the present data, its prominence is in part a function of the design. Teachers were asked to nominate one student who was experiencing significant behavioral problems, and overt disruptive behavior was the most obvious reason for nomination. Thus, the size of Factor 1 should not be interpreted as an indication of the relative prevalence of externalizing difficulties in the general school population.

This lack of significant difference between gender in total score ($t(122) = 1.29$, $p = .199$, $d = 0.24$, 95% CI for the mean difference (-0.124, 0.588)) differs from the results of Western normative samples, where boys tend to have higher behavioral difficulty scores than girls do (Achenbach & Rescorla, 2001; Goodman, 2001). One could be tempted to attribute the absence of a gender difference to a culturally specific rating pattern in non-Western contexts. There is, however, no regional evidence to support that interpretation. In the largest Saudi normative study of rated child behavior to date, Pierce et al. (2025) examined teacher ratings for 2,595 students and parent ratings for 3,127 youth and reported that boys were rated as having more symptoms than girls were but that measurement invariance was supported across gender groups. In line with this, Alqahtani (2010), who conducted a study with Saudi school children in grades 1 to 3, reported that boys tended to score higher on symptoms than girls did. It is thus clear that gender differences in teachers' ratings of behavior are measurable for Saudi samples and that a general attenuation of the gender effect outside of Western contexts is unlikely to explain the null effect demonstrated in this study.

The significant educational level effect ($F(3, 120) = 7.95$, $p < .001$) and the post hoc finding of the highest behavioral scores at the kindergarten level are broadly consistent with international longitudinal data suggesting that many behavioral difficulties peak in early childhood and are partially reversed with developmental maturation (Barkley, 2015; Faraone et al., 2021). The modest increase observed at the high school level may reflect the emergence of behavioral difficulties associated with adolescent psychopathology. This encompasses mood disorders and motivational disengagement, as reported in developmental psychopathology (Mash & Wolfe, 2019). This interpretation should be viewed with some skepticism; however, as the high school group included only 12 cases, their mean was not significantly different from that of the other groups, primary and middle school. Thus, the apparent effect may be due to sampling variance and not a developmental effect. The overall low R2 in the regression model

(6.4%) explains why the demographic variables measured here are not able to explain the majority of the variance in the behavioral scores. This is not, however, to be interpreted as an indication of the generalizability of the scale. A weak association with the demographic covariates suggests that these specific factors are not good predictors of the overall score. Generalizability across groups is an issue of measurement invariance, which is beyond the scope of the present sample because of its small size and unequal division across levels.

The three-factor structure has immediate implications for planning school-based assessment and intervention. The Externalizing Behavioral Dysregulation subscale may be used as a rapid screening tool to identify students who may have conduct problems and disruptive behavioral disorders for subsequent referral to specialists for more comprehensive assessment. The Attentional Dysregulation subscale is helpful for generating actionable information for classroom accommodations (preferential seating, task chunking, executive function scaffolding, etc.) that align with best practice guidelines for working with children with AD/HD. The Social Withdrawal and Anxiety subscale is used to identify students who do not have visible symptoms but whose internalizing difficulties are no less impaired in terms of their academic engagement and social development.

Within a multitiered system of supports (MTSS) or response to intervention (RTI) framework, scores and subscale profiles on the entire scale may be used to triage students into Tier 1 (universal monitoring), Tier 2 (group support), or Tier 3 (intensive individualized intervention) pathways (Sugai & Horner, 2009). The scale is also useful for progress monitoring so that practitioners can monitor behavioral change over periods of intervention. However, it must be emphasized that the instrument should not be used in isolation to arrive at any diagnostic conclusions. It should be triangulated through the use of structured clinical interviews, direct observation, achievement records, and, where available, clinical developmental evaluation.

CONCLUSION

Thirty teacher-rated behavioral items were developed and underwent initial psychometric validation in the current study to assess behaviorally stable patterns in school-aged children linked to the concepts of behavioral genetics theory. The scale cannot measure genetic variation, and there is no evidence of genetic causation in this study. A three-factor solution of externalizing behavioral dysregulation, attention dysregulation, and social withdrawal and anxiety was supported by the results of exploratory factor analysis. All three factors demonstrated excellent reliability, with Cronbach's alpha values exceeding 0.88, and the full scale achieved an overall alpha of 0.959. KMO = 0.92 and the cumulative variance explained = 69.55%, suggesting that the factor structure is well determined. Parallel analysis, however, confirmed the first two factors without any ambiguity and the third factor at the decision boundary; therefore, a three-factor solution may be considered provisional until the results are confirmed in an independent sample.

Behavioral scores differed significantly across educational levels; thus, kindergarten students had the highest scores, which is consistent with developmental accounts of early behavioral expression. Gender differences were nonsignificant, and education was the only significant demographic predictor of the regression model. The results offer preliminary evidence for the continued development and validation of the instrument for behavior screening in schools in Saudi Arabia. This is a limited claim. The evidence used here is a convenience sample of 124 teacher ratings in Riyadh only and obtained through an exploratory factor analysis. The content of the findings has therefore not been transferable to other Arabic-speaking or Gulf-region populations, whose dialect, curriculum and classroom norms vary, and to these populations, independent validity testing is needed in each context. The current data do not support the use of the instrument for individual intervention planning or for behavioral phenotype research. This essentially requires evidence of typical validity, such as confirmatory factor analysis, measurement invariance, and age-referenced norms. It may be considered a developing research instrument for use in the context of Saudi schools instead of an instrument that is validated for actual use in decision-making for individual children. The scale is most effectively used as a component of a tiered support system as one element of a comprehensive assessment protocol and is triangulated with clinical interviews, observations, and review of records. Future research should establish a three-factor structure with confirmatory factor analysis and measurement invariance and create age-referenced normative data.

Research Implications, Limitations and Future Study Directions

Theoretical Implications

The three-factor solution found in this study adds to behavioral genetics theory by showing that teacher-rated behavioral phenotypes organized themselves into theoretically coherent and empirically distinct dimensions. The convergence of the factor structure and the known externalizing, attentional and internalizing dimensions provides support for the validity of dimensional psychometric models in measuring genetically influenced behavioral traits in an applied context. The significant interfactor correlations indicate that these dimensions share etiological variance, which is consistent with the general factor of psychopathology identified in multivariate behavioral genetic studies (Polderman et al., 2015). Future molecular genetic studies could examine whether the factor scores derived from this instrument show the heritability patterns predicted by behavioral genetics theory.

Practical Implications

The scale offers an efficient, quick, low-cost screening tool that does not require specialist training for its administration for educational practitioners. This instrument of teacher ratings can be used in conjunction with school-based assessment procedures and can be helpful in the early identification of children who have significant behavioral needs. The instrument is a school-specific behavioral profile for clinical psychologists to complement parent-reported and clinician-administered instruments. However, the striking results of the high rates of behavioral scores at kindergarten may provide an argument for including the scale in the regular screening of young

children in the Saudi educational administration. Early identification represents the greatest window of opportunity for preventive intervention during the kindergarten years.

Ethical and Cultural Considerations

Interpreting genetically influenced traits in school settings requires careful attention to the risk of deterministic narratives. The present instrument captures behavioral stability and resistance to standard intervention; it does not provide evidence of genetic causation. The results should be framed as risk indicators that inform supportive environmental design, not as immutable labels. Cross-cultural measurement invariance should be formally established before high-stakes comparative use across different national contexts, gender groups, or age ranges. Communication with caregivers and teachers about the scope, limitations, and next steps following scale administration is essential to the ethical use of this instrument.

Limitations

Some drawbacks of the current study should be noted. First, quantitative testing of content validity was not carried out. The item pool was reviewed by 5 experts; however, there was no computation of a content validity ratio (CVR) or content validity index (I-CVI, S-CVI), and there was no numerical criterion for when an item was kept, revised, or removed. Future revision of the instrument should start with quantitative content validity ratings (CVR) from an expert panel, with the CVR and CVI values for each item reported with a set retention value. Second, there was no evidence of convergent or discriminant validity. Scale scores were not correlated with existing measures of the constructs or tested against unrelated constructs. Construct validity is thus based exclusively upon internal structure and theoretical consistency, and the interpretation of the subscale scores is tentative. This gap directly limits the practical applicability of the scale across school levels, which is key to the study's claims. Third, the sample consisted only of students from Riyadh, and the students who were rated were focal cases identified as having behavioral problems. The sample is thus a behaviorally referred group and is not representative of the general student population. This limits the generalizability of its results to other parts of Saudi Arabia and, to a greater extent, to other Arabic-speaking populations. Fourth, the behavioral ratings were made by a single informant type. Parent and clinician ratings provide a more complete picture of behavior functioning across contexts through multi-informant designs. Fifth, the label of genetically influenced behavior is psychometric, not molecular or neurobiological, and no genetic or neurobiological data were collected, nor is there direct evidence of genetic causation.

Future research directions

Several future directions for research are suggested. First, a confirmatory factor analysis with an independent sample should be performed to test the fit of the three-factor model and to compare the three-factor model with alternative models such as a higher-order model and a bifactor model. Second, high-stakes comparative use requires measurement invariance testing across gender, age groups, and cultural contexts. Third, temporal stability should be assessed with test-retest reliability studies that have two- to four-week intervals. Fourth, convergent and

discriminant validity should be established by correlating scale scores to established measures of AD/HD, anxiety, and conduct problems. Fifth, criterion-related validity should be examined by comparing the scale scores of children who are referred for behavioral evaluation and those of children in samples from the general classroom, with ROC analyses used to determine the best screening cutoff points. Sixth, item response theory analyses are used to provide item-level psychometric information that is not provided by classical test theory, such as discrimination parameters and differential item functioning by sex and age. Seventh, the development of age-referenced normative data would substantially increase the clinical utility of the instrument for the interpretation of individual scores.

Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available because of ethical considerations related to participant confidentiality but are available from the corresponding author upon reasonable request.

Conflict of Interest Statement

The authors declare that they have no known competing financial or nonfinancial interests that could have appeared to influence the work reported in this paper.

REFERENCES

Abo Hamza, E., Weickert, T., Ali, M., & Moustafa, A. (2020). Reward and punishment learning in schizophrenia and bipolar disorder. *Behavioural Brain Research, 381*, 112298. https://doi.org/10.1016/j.bbr.2019.112298

Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry, 49*(3), 251–275. https://doi.org/10.1111/j.1469-7610.2007.01867.x

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. University of Vermont, Research Center for Children, Youth, and Families.

Alqahtani, M. M. J. (2010). Attention-deficit hyperactive disorder in school-aged children in Saudi Arabia. *European Journal of Pediatrics, 169*(9), 1113–1117. https://doi.org/10.1007/s00431-010-1190-y

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., ... Neale, B. M. (2018). Analysis of shared heritability in common disorders of the brain. *Science, 360*(6395), eaap8757. https://doi.org/10.1126/science.aap8757

Bandura, A. (1977). *Social learning theory*. Prentice-Hall.

Barkley, R. A. (2015). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th ed.). Guilford Press.

Beyazıt, U., Yıldız Akkuş, S., & Bütün Ayhan, A. (2023). Predictor Role of Parental Neglect On Socio-Emotional And Moral Development And Behavioral Problems In Children. *Theory and Practice in Child Development*, *3*(2), 98–118. https://doi.org/10.46303/tpicd.2023.14

Conners, C. K. (2008). *Conners 3rd edition manual*. Multi-Health Systems.

Cortese, S., & Coghill, D. (2018). Twenty years of research on attention-deficit/hyperactivity disorder: Looking back, looking forward. *Evidence-Based Mental Health, 21*(4), 173–176. https://doi.org/10.1136/ebmental-2018-300050

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10, Article* 7. https://doi.org/10.7275/jyj1-4868

Demontis, D., Walters, G. B., Athanasiadis, G., Walters, R., Therrien, K., Nielsen, T. T., Farajzadeh, L., Voloudakis, G., Bendl, J., Zeng, B., Zhang, W., Grove, J., Als, T. D., Duan, J., Satterstrom, F. K., Bybjerg-Grauholm, J., Bækved-Hansen, M., Gudmundsson, O. O., Magnusson, S. H., ... Børglum, A. D. (2023). Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics*, 55(2), 198–208. https://doi.org/10.1038/s41588-022-01285-8

DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). SAGE Publications.

Eapen, V., & Ghubash, R. (2004). Help-seeking for mental health problems of children: Preferences and attitudes in the United Arab Emirates. *Psychological Reports, 94*(2), 663–667. https://doi.org/10.2466/pr0.94.2.663-667

Elsantil, Y., & Abo Hamza, E. (2021). A review of internal and external factors underlying the purchase of counterfeit products. *Academy of Strategic Management Journal, 20*(1), 1–13. https://www.abacademies.org/articles/A-review-of-internal-and-external-factors-underlying-the-purchase-of-counterfeit-products-1939-6104-20-1-689.pdf

Emam, M. M., Hilal, M. M. A., Kazem, A. M., & Alkharousi, S. J. (2016). Psychometric properties of the Arabic self-report version of the strengths and difficulties questionnaire. *Research in Developmental Disabilities,* 59, 211–220.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., Newcorn, J. H., Gignac, M., Al Saud, N. M., Manor, I., Rohde, L. A., Yang, L., Cortese, S., Almagor, D., Stein, M. A., Albatti, T. H., Aljoudi, H. F., Alqahtani, M. M. J., Asherson, P., ... Wang, Y. (2021). The World Federation of ADHD International Consensus Statement: 208 evidence-based conclusions about the disorder. *Neuroscience & Biobehavioral Reviews, 128*, 789–818. https://doi.org/10.1016/j.neubiorev.2021.01.022

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*(1), 93–103. https://doi.org/10.1177/014662168701100107

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

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*(4), 636–645. https://doi.org/10.1176/appi.ajp.160.4.636

Jami, E. S., Hammerschlag, A. R., Ip, H. F., Allegrini, A. G., Benyamin, B., Border, R., Diemer, E. W., Jiang, C., Karhunen, V., Lu, Y., Lu, Q., Mallard, T. T., Mishra, P. P., Nolte, I. M., Palviainen, T., Peterson, R. E., Sallis, H. M., Shabalin, A. A., Tate, A. E., ... Middeldorp, C. M. (2022). Genome-wide association meta-analysis of childhood and adolescent internalizing symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry*, 61(7), 934–945. https://doi.org/10.1016/j.jaac.2021.11.035

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Mash, E. J., & Wolfe, D. A. (2019). *Abnormal child psychology* (7th ed.). Cengage Learning.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moffitt, T. E., Arseneault, L., Jaffee, S. R., Kim-Cohen, J., Koenen, K. C., Odgers, C. L., Slutske, W. S., & Viding, E. (2008). Research review: DSM-V conduct disorder: Research needs for an evidence base. *Journal of Child Psychology and Psychiatry, 49*(1), 3–33. https://doi.org/10.1111/j.1469-7610.2007.01823.x

Moore, D. S., & Shenk, D. (2017). The heritability fallacy. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1–2), e1400. https://doi.org/10.1002/wcs.1400

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry, 58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Perold, M., Louw, C., & Kleynhans, S. (2010). Primary school teachers' knowledge and misperceptions of attention deficit hyperactivity disorder (ADHD). *South African Journal of Education, 30*(3), 457–473.

Pierce, C. D., Alhossein, A., Lambert, M. C., Aldabas, R., Bakhiet, S. F. A., & Fitzpatrick, H. (2025). ADHD rating scale adaptation in Saudi Arabia: Factor structure, measurement

invariance, and normative data. *BMC Psychology, 13*, Article 1101. https://doi.org/10.1186/s40359-025-03398-1

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 findings from behavioral genetics. *Perspectives on Psychological Science, 11*(1), 3–23. https://doi.org/10.1177/1745691615617439

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics, 47*(7), 702–709. https://doi.org/10.1038/ng.3285

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*(1), 13–43. https://doi.org/10.1207/S15328031US0201_02

Reynolds, C. R., & Kamphaus, R. W. (2015). *BASC-3: Behavior Assessment System for Children* (3rd ed.). Pearson.

Rutter, M. (2006). *Genes and behavior: Nature-nurture interplay explained*. Blackwell Publishing.

Samarakkody, D. C., Fernando, D. N., Perera, H., McClure, R. J., & De Silva, H. (2010). The Child Behaviour Assessment Instrument: Development and validation of a measure to screen for externalising child behavioural problems in community setting. *International Journal of Mental Health Systems,* 4, Article 13. https://doi.org/10.1186/1752-4458-4-13

Scaini, S., Belotti, R., & Ogliari, A. (2014). Genetic and environmental contributions to social anxiety across different ages: A meta-analytic approach to twin data. *Journal of Anxiety Disorders, 28*(7), 650–656. https://doi.org/10.1016/j.janxdis.2014.07.002

Sciutto, M. J., Terjesen, M. D., & Frank, A. S. B. (2000). Teachers' knowledge and misperceptions of attention-deficit/hyperactivity disorder. *Psychology in the Schools, 37*(2), 115–122. https://doi.org/10.1002/(SICI)1520-6807(200003)37:2<115::AID-PITS3>3.0.CO;2-5

Seleem, M. A., Amer, R. A., Elhosary, M., Saada, S., Abo Hamza, E., Elfert, Y., Abd El-fatah Abdo, S., Kabbash, I., & Achenbach, T. M. (2023). Psychometric properties and cross-cultural comparison of the Arabic version of the Child Behavior Checklist (CBCL), Youth Self Report (YSR), and Teacher's Report Form (TRF) in a sample of Egyptian children. *Middle East Current Psychiatry,* 30, Article 61. https://doi.org/10.1186/s43045-023-00328-y

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality, 17*(4), 223–237. https://doi.org/10.1080/09362830903235375

Supena A., Kusmawati A.P. (2022). Identification of student with attention deficit hyperactivity disorder (ADHD): Experiences of primary school teacher's assessment in inclusive

school. *Journal of Social Studies Education Research, 13* (2), 55 – 82. https://jsser.org/index.php/jsser/article/view/3985/562

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9(5), 160–164. https://doi.org/10.1111/1467-8721.00084

World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th rev.). https://icd.who.int/

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Zelazo, P. D., & Carlson, S. M. (2020). The neurodevelopment of executive function skills: Implications for academic achievement gaps. *Psychology & Neuroscience, 13*(3), 273–298. https://doi.org/10.1037/pne0000208

APPENDIX

Table 2.

Descriptive Statistics for All 30 Scale Items (N = 124)

#	Item Description	M	SD	Skew	Kurt
1	Shows rapid reactions without a clear stimulus	2.51	1.46	0.43	-1.23
2	Avoids social interaction despite repeated attempts	2.48	1.44	0.47	-1.12
3	Shows anxiety or social withdrawal since early childhood	2.15	1.29	0.71	-0.75
4	Tends toward persistent isolation from peers	2.11	1.44	0.93	-0.61
5	Displays stereotyped behaviors that are difficult to modify	2.32	1.44	0.63	-1.04
6	Shows persistent hyperactivity not explained by the environment	2.27	1.45	0.71	-0.93
7	Noticeably avoids eye contact with others	2.15	1.27	0.80	-0.42
8	Shows hypersensitivity to sounds or lights	1.96	1.27	1.21	0.36
9	Behavior is unpredictable in simple social situations	2.35	1.43	0.53	-1.11
10	Shows repetitive behaviors without a clear purpose	2.35	1.47	0.55	-1.20
11	Shows clear difficulties in sustaining attention	2.50	1.35	0.43	-1.02
12	Shows internal tension without clear reasons	2.37	1.39	0.53	-1.05
13	Speaks negatively about own abilities and competence	2.07	1.33	1.00	-0.25
14	Lacks enthusiasm and motivation for learning	2.54	1.43	0.45	-1.12
15	Has difficulty concentrating despite suitable environment	2.54	1.39	0.34	-1.22

#	Item Description	M	SD	Skew	Kurt
16	Hesitates frequently before carrying out assigned tasks	2.52	1.38	0.43	-1.13
17	Avoids initiating activities despite being familiar with them	2.48	1.41	0.54	-0.99
18	Shows mental distraction during instruction	2.69	1.33	0.22	-1.14
19	More preoccupied with internal thoughts than surroundings	2.60	1.44	0.39	-1.16
20	Shows slow responses when exposed to learning situations	2.58	1.34	0.44	-0.93
21	Does not participate in activities despite available support	2.61	1.42	0.44	-1.08
22	Repeats inappropriate behaviors in the classroom	2.48	1.44	0.45	-1.16
23	Imitates negative behaviors observed in peers	2.52	1.43	0.43	-1.19
24	Refuses to follow classroom instructions	2.35	1.39	0.60	-0.93
25	Makes noticeable sounds or movements during class	2.50	1.55	0.45	-1.34
26	Withdraws repeatedly from learning situations	2.44	1.36	0.54	-0.91
27	Deliberately provokes peers during learning activities	2.34	1.52	0.70	-0.99
28	Talks during instruction without permission	2.58	1.50	0.35	-1.32
29	Moves around the classroom without a clear reason	2.29	1.55	0.74	-1.02
30	Responds negatively to instructional guidance	2.23	1.37	0.77	-0.65

Note. M = mean; SD = standard deviation; Skew = skewness; Kurt = excess kurtosis. All the values are based on N = 124.

Table 4.

Promax-Rotated Pattern Matrix: Three-Factor Solution (N = 124)

Item Description	F1 Beh. Dysreg.	F2 Atten. Dysreg.	F3 Soc. W/A	h^2
Shows rapid reactions (Item 1)	0.617*	0.037	0.102	0.392
Avoids social interaction (Item 2)	0.060	0.462*	0.494*	0.461
Anxiety/social withdrawal (Item 3)	0.098	0.383*	0.710*	0.661
Persistent isolation (Item 4)	0.032	0.406*	0.708*	0.667
Stereotyped behaviors (Item 5)	0.693*	0.180	0.342	0.630
Persistent hyperactivity (Item 6)	0.734*	0.166	0.232	0.620

Item Description	F1 Beh. Dysreg.	F2 Atten. Dysreg.	F3 Soc. W/A	h^2
Avoids eye contact (Item 7)	0.125	0.488*	0.639*	0.662
Hypersensitivity to stimuli (Item 8)	0.250	0.210	0.603*	0.470
Unpredictable behavior (Item 9)	0.701*	0.229	0.356*	0.671
Repetitive behaviors (Item 10)	0.791*	0.136	0.297	0.732
Attention difficulties (Item 11)	0.319	0.618*	0.232	0.538
Internal tension (Item 12)	0.334	0.473*	0.501*	0.587
Speaks negatively about self (Item 13)	0.283	0.324	0.460*	0.397
Lacks enthusiasm (Item 14)	0.139	0.789*	0.224	0.692
Difficulty concentrating (Item 15)	0.258	0.846*	0.115	0.796
Hesitates before tasks (Item 16)	0.141	0.841*	0.200	0.767
Avoids initiating activities (Item 17)	0.202	0.802*	0.213	0.729
Mental distraction in class (Item 18)	0.410*	0.654*	0.091	0.604
Internal preoccupation (Item 19)	0.226	0.766*	0.200	0.677
Slow responses to learning (Item 20)	0.104	0.861*	0.158	0.778
Does not participate (Item 21)	0.180	0.739*	0.282	0.658
Repeats inappropriate behaviors (Item 22)	0.880*	0.186	0.055	0.812
Imitates negative peers (Item 23)	0.757*	0.241	-0.084	0.637
Refuses instructions (Item 24)	0.785*	0.300	0.144	0.728
Makes noises/movements (Item 25)	0.896*	0.062	0.124	0.823
Withdraws from learning (Item 26)	0.560*	0.541*	0.201	0.646
Deliberately provokes peers (Item 27)	0.884*	0.155	-0.089	0.813
Talks without permission (Item 28)	0.871*	0.160	0.023	0.786
Moves without reason (Item 29)	0.829*	0.092	-0.014	0.696
Responds negatively to guidance (Item 30)	0.798*	0.320	0.087	0.747

Note. F1 = Externalizing Behavioral Dysregulation (14 items); F2 = Attentional Dysregulation (9 items); F3 = Social Withdrawal and Anxiety (7 items). Beh. Dysreg. = Externalizing Behavioral Dysregulation; Atten. Dysreg. = attention dysregulation; Soc. W/A = social withdrawal and anxiety. h^2 = communality. Loadings >= |0.35| are asterisked (*) and considered meaningful. Extraction: Principal Axis Factoring; Rotation: Promax (oblique).

Table 7.

Corrected Item-Total Correlations and Alpha-if-Item-Deleted Values

Item	Description (Abbreviated)	CITC	alpha-if-Deleted
1	Shows rapid reactions	.434	.960
2	Avoids social interaction	.402	.960
3	Anxiety or social withdrawal	.544	.959
4	Persistent isolation	.506	.959
5	Stereotyped behaviors	.709	.957
6	Persistent hyperactivity	.688	.958
7	Avoids eye contact	.600	.958
8	Hypersensitivity to stimuli	.507	.959
9	Unpredictable behavior	.749	.957
10	Repetitive behaviors	.728	.957
11	Clear attention difficulties	.672	.958
12	Internal tension	.694	.958
13	Speaks negatively about abilities	.555	.959
14	Lacks enthusiasm	.638	.958
15	Difficulty concentrating	.717	.957
16	Hesitates before tasks	.661	.958
17	Avoids initiating activities	.687	.958
18	Mental distraction	.707	.957
19	Internal preoccupation	.679	.958
20	Slow responses	.631	.958
21	Does not participate	.660	.958
22	Repeats inappropriate behaviors	.729	.957
23	Imitates negative peers	.626	.958
24	Refuses instructions	.772	.957
25	Makes noises or movements	.687	.958
26	Withdraws from learning	.786	.957
27	Deliberately provokes peers	.655	.958
28	Talks without permission	.694	.958
29	Moves around classroom	.606	.958
30	Responds negatively to guidance	.771	.957

Note. CITC = corrected item–total correlation. The alpha-if-deleted column shows the Cronbach's alpha that would result if that item were removed from the full 30-item scale.