A National School-based Screening Program for Mental Health Problems Among Children aged 6 to 12 years in Iran: Scale Development and Psychometric Evaluation

Saeed Akbari Zardkhaneh[1], Mohsen Rajabi[2,3], Ali Mohammad Zanganeh[4], Saeed Namaki[5], Nader Mansour Kiaee[6], Mohsen Jallalat Danesh[4], and Siyamak Tahmasebi Garmatani[7]

[1] Department of Developmental Psychology, Shahid Beheshti University, Tehran, Iran

[2] Department of Psychology, University of Tehran, Tehran, Iran

[3] Institute for Connected Communities (ICC), University of East London, London, UK

[4] Preventive Department of Judicial System, Tehran, Iran

[5] Department of Immunology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[6] Social Communications Sciences, Faculty of Social Sciences, Allameh Tabataba'i University, Tehran, Iran

[7] Department of Preschool Education, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

Abstract

Schools are an ideal setting in which to measure and promote mental health difficulties. The aim of present study was to develop the Nemad Electronic Mental-Health Assessment Devices for Children (NEMAD-C) aged 6 to 12 years in Iran. A sample of parents and teachers (N = 10,163) were recruited to complete the parent and teacher reports. Totally, explorative and confirmatory factor analyses showed that the eight-factor model provides a better fit for both parental report and teacher report versions. Results revealed a screening tool consisting of eight dimensions: child abuse risk, self-harm, anxiety, depression, disruptive behavior disorders, attention deficit/hyperactivity disorders, academic achievement deficit, and self-regulation. Findings showed that the internal consistency coefficients of the subscales were high, and convergent validity was evidenced by significant correlations with theoretically related constructs. Therefore, the NEMAD-C has adequate reliability and validity and could be used for multi-dimensional assessment of mental health problems in Iran.

***Keywords*:** Iranian students mental health, school mental health, electronic assessment tool, scale development, screening tool, school-based screening, NEMAD project

Word count: 7636

Introduction

Mental disorders cause undesirable outcomes in children, affecting them in many areas including educational performance, social interactions, physical problems, and unwanted costs for societies (Haile et al., 2017; Trautmann et al., 2016). Furthermore, childhood psychological difficulties impact their future mental health in adulthood (Newbury et al., 2018). Effective schoolwide screening process contributes to identify children who will benefit from appropriate intervention in earlier stages to create improved learning environment (Kettler et al., 2014). As early educational and emotional/behavioral problems may result in long-term and ongoing issues, it is worth noting that efficient screening is conducted in the early stages of childhood.

Current methods of early identification of mental health problems in psychoeducational settings vary widely. These approaches comprise parent referral, teacher referral, pupil self-referral through school-based or community-based services, and universal assessment tools (Eklund et al., 2009). Studies indicate that teachers do not refer students based on psychological and social problems at the same rate as other academic concerns (Walker et al., 2000). Moreover, many teachers and school staff assume student's psycho-social problems as some else's responsibility (Severson et al., 2007) which leads to lower referral rates. Regarding this, Eklund et al. (2009) suggested that teachers can only identify 23 percent of children with behavioral and emotional problems.

In response to the need to screen students' mental health, a number of universal screening tools have been developed and validated. These instruments include Social Skills Improvement System (SSIS; Elliot & Gresham, 2008), the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001), Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992), and Student Risk Screening Scale (SRSS; Drummond, 1994). Many of these tools lack cross-cultural equivalence and may not be adequately fit to measure and screen mental health problems from non-Western cultures.

Reviewing current screening tools showed following limitations. First, screeners in use do not include self-harm behaviors that are thought to be common among children. As such, children's self-harm behaviors may go undetected. Deliberate self-harm (DSH) is one of the strongest predictors of suicide (Hawton et al., 2015) which is the second leading cause of death among 10 to 24-year olds worldwide, accounting for 6.3% of all deaths (Patton et al., 2009). Suicide is the third leading cause of death among youth in low- and middle-income countries (LMIC), accounting for 8% of all deaths among 15- to 29-year olds (Whiteford et al., 2015). There is little information worldwide about suicidal behavior in children under the age of ten (Miller, 2021). Across

cultures, there is no standardized culture-specific or psycho-social assessment tool in school settings to assess the immediate risk of self-harm for children.

Second, self-regulation, which is a fundamental aspect of early childhood development, has been overlooked in many social and behavioral screening tools (McCoy, 2019). Self-regulation covers a wide range of essential qualities, including the ability to recognize others' emotions, the capability to have social and dynamic interactions with others, the capacity to manage and control aggressive and violent behavior, and development of attention, working memory and inhibitory control (McClelland et al., 2010). Cognitive constructs that contribute to the self-regulation are often defined as executive functions and they encompass a system of diverse cognitive skills which enable the individual to flexibly adapt to changes and unstable situations; switch between tasks; update and integrate information in the working memory; perform self-oriented actions; and inhibit automatic or responses to pursue a goal (Hughes & Ensor, 2011). In recent years, there has been growing evidence that early assessment of self-regulation predicts later mental and physical health, educational achievements and family social risks in children (Woodward et al., 2017; Arslan, 2018). Higher levels of self-regulation have been associated with desirable outcomes on academic attainment, school readiness skills and attendance, physical health, and mental health (Woodward et al., 2017). In contrast, lower levels of self-regulation has been strongly related to negative consequences in children, such as developmental impairments, behavior problems, psychological disorders, drugs abuse/addiction, and unemployment (Palacios-Barrios & Hanson, 2019; Woodward et al., 2017).

Third, early assessment of adverse childhood experiences (ACEs) has been recommended as an effective way to prevent child abuse, maltreatment and neglect (Klevens & Whitaker, 2007). A number of standardized tests are used to predict future risk of ACEs and child abuse such as The Maternal History Interview 2 (Brayden et al., 1993), the Kempe Family Stress Inventory (Peters & Barlow, 2003), Family Psychosocial Risk Inventory (Hunter et al., 1978), Dunedin Family Services Indicator (Muir et al., 1989), and the Child Abuse Potential Inventory (CAPI; Milner et al., 1984), but child abuse has not been embedded in screening tools.

Forth, an important reason that universal screening has not been properly integrated into wider service delivery, in educational settings or schools in particular, is that there exists a shortage of detailed and operating instructions that enable universal screening to be performed in a practical way with high levels of fidelity. Verlenden et al. (2020) recently found that common obstacles to the implementation of universal screening are being time-consuming and costing a lot of materials and human resources. This process can be facilitated through implementing electronic systems since many schools and families may have difficulties in accessing to

centers serving clinical measures (WHO, 2006). For this purpose, we developed an electronic system aimed to screen and monitor mental health conditions in adolescents who are at risk of psychological and social problems. Electronic systems possess a variety of advantageous in spite of outpatient visits in which clients should pay heavy costs for receiving mental health services and they are not easily available in all areas (Andrews et al., 2000). Many young may hesitate to refer to mental health services because of social isolation and stigma (Stephens-Reicher et al., 2011). Using online and electronic assessment systems can surmount fiscal, geographic and psychological barriers (Stephens-Reicher et al., 2011).

Although mental health research in Asian and Middle Eastern countries has gained increasing attention in recent years, there is currently a regrettable lack of validated and culturally appropriate scales and screening tools with adequate psychometric properties (Kaiser et al., 2019). Also, widespread methodological concerns regarding the cross-cultural validity and cultural adaptation, continue to overshadow this research field (e.g., Maalouf et al., 2019). Further considerations should be given to whether the administration approaches and identification outcomes associated with a screening assessment are compatible with the local and community-based services delivery needs (Glover & Albers, 2007). In addition, western conceptualizations of mental health have been found not to be universally valid (Patel, 2001) and it is now more or less generally assumed that mental health cannot be assessed in Middle Eastern contexts without bringing local views into account (Patel et al., 2008). Although a number of mentioned instruments have adequate psychometric characteristics, many possess a number of limitations. Specifically, many of these instruments lack national norms, and do not cover the cultural diversity which should be included into the screening systems and prevent from misidentifying normal behaviors in one culture as abnormal (Dowdy et al., 2014). Current screening tools are not based on Iranian culture. This process should be done with caution, as there are fundamental cultural and religious differences between the different societies (Loewenthal, 2019; Kirmayer et al., 2017). Such differences, if not taken into account, may affect the reliability and validity of instruments and produce faulty results (Mahmood et al., 2015). Although psychometric assessment of culture-specific tools is an active line of research in Iran (e.g., Akbari Zardkhaneh et al., 2021; Toosi et al., 2017), such measures for assessment of children's mental health are widely missing. In order to fill the dearth of large-scale research on Iranian children's mental health, developing a culturally appropriate scale which is based on Iranian culture and mental disorders' epidemiological prevalence rate is necessary, but lacking in the literature. Therefore, the present study aims to develop a culturally appropriate scale for elementary school students: the Nemad Electronic Mental-Health Assessment Devices (NEMAD). Nemad Electronic Mental-Health Assessment Devices (NEMAD) is part of a

national project for screening and promotion of mental health and besides that, it is a Persian acronym derived of initial letters of social care system for students. The present study aims to evaluate the validity and reliability of the NEMAD for children (NEMAD-C).

Materials and Methods

Participants

Symptom ratings for children were obtained from parents and teachers of elementary school students residing in 5 provinces of Iran. Parents and teachers were instructed to complete a demographic history form and one scale for each child between the ages of 6 to 12 years old. Parents were also requested to grant permission to contact the child's teacher and obtain school ratings. Participation of all parents and teachers was voluntary. Information was gathered from 10,163 children. The mean age of the sample group was 9.83 (SD= 4.33) years, with females accounting for 52% of the sample.

In 7,525 of parents, 76% were females. 44% of the mothers and 56% of the fathers have completed high school of equivalent, while only 21% of the mothers and 32% of the fathers have completed college on equivalent. Religious breakdown of the families revealed approximately 87% Shia, 12% Sunni, 1% others. In 2,638 of teachers, 86% were females, 82% of them BA and others MA and higher. Religious breakdown of the families revealed approximately 91% Shia, 8.5% Sunni, 0.5% others.

Procedures

Devellis (2012) suggests that construct validity is the first step in measure development. Construct validity is constructed via substantive and structural validity evidences. Substantive validity consists of two processes: determining the nature and scope of the construct of interest, and creating an item structure and item pool. Structural validity is established through evaluating the latent structure, internal consistency, and concurrent of predictive correlational analyses.

At the initial step in measure development process, we determined the nature and scope of measure we intended to develop: A measure for screening children's mental health based on Iranian culture. Next, we conceptualized the nature of the construct to be assessed by the measure. Since a strong theoretical model of common mental health problems among children aged 7 to 11 was required as a basis for the instrument (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), a comprehensive literature review of child mental health, including

definitions, identifying existing instruments, and prevalence studies in primary care and clinical populations was conducted. Results showed that, 20 problems are more prevalent in school aged children. In order to finalize the model, twenty specialists in the area of child and adolescent psychology were invited to extract the model. The final model compromised of eight components based on the most prevalent problems: Anxiety (ANX), Depression (DEP), Attention Deficit/ Hyperactivity Disorder (ADHD), Disruptive Behavior (DB), Self-Harm risk (SH), Self-Regulation problems (SR), Child Abuse Risk (CAR), Academic Achievement Deficit (AAD). After extracting the components, each components' indicator was extracted. Each indicator turned into statements. In this part, almost 200 items were created for Parent-Report (PR) and 150 items for Teacher-Report (TR). Then, expert panel group decided to omit 29 and 22 items in PR and TR respectively.

In the next step, a pool of items (171 items for PR and 128 Items for TR) related to each of the eight subscales were acquired. A group of experts, who were specialist in psychometric testing, clinical child psychology, educational psychology, social workers, and social skills teachers, critically assessed the items. They decided to omit 20 and 27 items for PR and TR, respectively. Therefore, the preliminary version of the NEMAD-C was developed with 150 items for PR and 103 items for TR forms.

A 4-point Likert-type response scale anchored only at the endpoints: 1 = totally agree, 2 = agree, 3 = disagree, and 4 = totally disagree, was used to obtain an interval level of measurement. In order to prevent response biases, 10 items in PR and 8 items in TR were coded reversely. Discussion groups (face to face and each group up to 20 people) were formed to examine the face validity of the preliminary version of NEMAD-C. A total of 600 teachers and 2,100 parents were asked to provide their qualitative comments about the fluency and readability of the items. Their comments led to changes in the structure and wording of some items. These samples were randomly selected from six provinces (West Azarbayjan, Alborz, Kermanshah, Sistan Balouchestan, Razavi Khorasan, Fars) with different culture in Iran.

After establishing the substantive validity of the NEMAD-C, the structural validity of the measure was tested. The structural validity of the NEMAD-C was tested in two phases. At the first phase, 6,971 parents and 2,419 teachers completed the forms. Each data file divided into two equal size groups for exploratory and confirmatory factor analyses. At the second phase, 554 parents and 219 teachers were asked to complete Child Behavior Checklist (CBCL; Achenbach, 1991a) and Teacher Report Form (TRF; Achenbach, 1991b) as the concurrent validity indexes in addition to the NEMAD-C-PR and NEMAD-C-TR. All measures were completed online via a national system of educational screening. There were no missing values because electronic devices were used for data gathering and participants were not allowed to leave an item unanswered.

Measures

Child Behavior Checklist (CBCL; Achenbach, 1991a). This is a 113-item parental questionnaire assessing child behavior problems, emotional difficulties, externalizing and internalizing symptoms. The CBCL is comprised of eight empirically based syndrome scales: Aggressive Behavior (AB), Anxious/Depressed (AD), Attention Problems (AP), Rule-Breaking Behavior (RB), Somatic Complaints (SC), Social Problems (SP), Thought Problems (TP), and Withdrawn/Depressed (WD). It is a well-known, worldwide used dimensional rating scale and psychometric properties of CBCL have been reported in most countries (Shahrivar et al., 2017).

Teacher Report Form (TRF; Achenbach, 1991b). The TRF is a 118-item scale that is designed to be completed by teachers. The TRF is comprised of scales separated into two sections based on item content: social competence/adaptive functioning and behavior problems. Items are scored with a zero (*not true*), 1 (*sometimes true*), or 2 (*very true or often true*). The TRF have been standardized to obtain normative reference points (i.e., Achenbach & Edelbrock, 1983). As such, standard T scores quantify a student's standing in relation to other students and determine whether elevated scores on a particular scale falls in a clinical range. Psychometric properties of TRF have been well established in several countries (Shahrivar et al., 2017).

Results

Item analysis

Item analysis was conducted on the exploratory sub-sample. Analysis for each item included the calculation of eight criteria: Item Mean (IM), Standard Deviation (SD), Skewness (SK), Kurtosis (KU), Squared Multiple Correlation (SMC), Corrected Item Total Correlation (CITC), Chronbach's Alpha If Item Deleted (CAIID) (DeVellis, 2012). Descriptive statistics for some of the items are presented in Table 1. The subscale correlations were all greater than 0.30, suggesting that the items could discriminate among the different disorders (Nunnally & Bernstein, 1994). Item analysis results showed that 8 items in PR and no items in TR were omitted. **|*Table 1 near here*|**

Exploratory factor analysis

Fabrigar et al. (1999) argued that if data are normally-distributed, maximum likelihood is the best choice because it allows for the computation of a wide range of indexes and permits statistical significance testing for factor loadings, correlations among factors, and the computation of confidence intervals (Fabrigar et al., 199).

Since the data are normally distributed, maximum likelihood was used as the extraction method. Parallel analysis (Franklin et al., 1995) also was used for determining the number of principal components. The initial number of factors identified by Monte Carlo software were 5 factors for PR form and 7 factors for TR form.

Using maximum likelihood as the extraction method and promax as the rotation method, 5 transformed factors with salient pattern coefficients > 0.30 provided the best fitting and interpretable solution. In PR form, of the 42 items that loaded on factor 1, 28 items were theoretically written for Child Abuse Risk (CAR) and Self-Harm (SH). 13 of 26 items that loaded on factor 2, were written for Academic Achievement Deficit (AAD). 24 items that loaded on factor 3, were written for Disruptive Behavior disorders (DB) and ADHD. Of the 32 items that loaded on factor 4, 25 of the items were originally written for the Anxiety (ANX) and Depression (DEP). Overall, 8 of the 9 items that loaded on factor 5, were written for Self-Regulation (SR). Exploratory factor analysis (EFA) for TR and PR were conducted on calibration sample groups.

In TR form, maximum likelihood was used as the extraction method and promax was selected as rotation method. In TR, of the 23 items that loaded on factor 1, 16 the items were theoretically written for the Academic Achievement Deficit (AAD). 8 of the 17 items that loaded on factor 2, were written for Attention deficit/hyperactivity disorder (ADHD). 15 of the 16 items that loaded on factor 3, were originally written for Child Abuse Risk (CAR) and Self-Harm (SH). 10 of the 22 items that loaded on factor 4, were written for Depression (DEP). 7 of the 9 items that loaded on factor 5, were written for Self-Regulation (SR), 3 of the 4 items that loaded on factor 6, were written for Anxiety (ANX) and all of the items that loaded on factor 7, were written for Disruptive Behavior disorders (DB).

Confirmatory factor analysis

For acquiring model fit, following indexes were used: Incremental Fit Index (IFI) (Tanaka, 1993), Comparative Fit Index (CFI) (Bentler, 1990), Non-Normed Fit Index (NNFI) (Bentler & Bonett, 1980), Root Mean Square Error of Approximation (RMSEA) (Stieger, 1990), Root Mean Square Error of Approximation Confidence Interval (RMSEA 90%CI) (Hu & Bentler, 1999), and Standardized Root Mean Square Residual (SRMSR) (Hu & Bentler, 1999). Both confirmatory factor analyses (CFA) for TR and PR forms were conducted on confirmatory sub-sample.

Six extracted factors from exploratory factor analysis (EFA) and eight factors based on expert panel (EP) were examined for invariability of factor structure for PR (Table 2). Since sample size and model complexity can affect the interpretation of chi square statistics (Bearden et al., 1982), an incremental fit index has been proposed by Bentler and Bonett (1980) as an alternative overall goodness of fit index when the sample size is

large. Following recommendations of Chen (2007) for model fit changes, if CFI ≤ 0.10, RMSEA ≤ 0.015, and SRMR ≤ 0.030 for tests of factor loading invariance, the measurement invariance would be evidenced and confirmed. Overall, these results indicated that both models yielded close fits to the data; however, given that parsimony is one of the scientific criteria used for model selection, and considering that the eight-mental-health domains loaded significantly onto the expert panel (EP) form, we identified the EP Model as the preferred structure over the EFA Model. Similarly, the EP model of TR was chosen as a final structure. **[*Table 2 near here*]**

Internal consistency reliability of subscales

Cronbach's alpha was computed to the internal consistency as reliability coefficient. As it is presented in Table 3, in PR form, all Cronbach's alpha coefficients were above satisfactory levels and most were above 0.80 except for two of the eight factors, namely Self-Regulation (α = 0.65) and Child Abuse Risk (α = 0.73). Also, in TR form, all the estimates were above 0.80 (Table 4). In PR form, Disruptive Behaviors were highly correlated with ADHD (0.81), followed by Academic Achievement Deficit and ADHD (0.78), Depression and Anxiety (0.76), Depression and ADHD (0.75). Lowest correlations were found between Self-Regulation and Self-Harm (0.28) and Child Abuse Risk (0.39) (Table 3). In TR form, highest correlations were found between Disruptive Behaviors and ADHD (0.89), Depression and Anxiety (0.89), Child Abuse Risk and Self-harm (0.87). Academic Achievement Deficit and Self-Harm were less strongly correlated (0.44), followed by Self-Regulation and Self-Harm (0.42) (Table 4). **[*Table 3 and 4 near here*]**

Convergent validity

The correlations among the NEMAD-C and CBCL are presented in Table 5. In sum, all NEMAD-C-PR subscales were correlated to all CBCL subscales in a significant level ($p < 0.01$), which ranged between 0.21 (Child Abuse Risk with Rule-Breaking Behaviors) to 0.74 (Disruptive Behavior disorders with Aggressive Behaviors). On the other side, all of the 8 subscales of NEMAD-C-TR were also related to all 8 subscales of CBCL with correlations ranging from 0.14 (Self-Regulation with Anxious/Depressed, and Disruptive Behavior disorders with Withdrawn/Depressed) to 0.73 (Disruptive Behavior disorders with Aggressive Behavior). As expected, in TR forms, Disruptive Behavior disorders had strong correlations with Aggressive Behaviors (0.73) and Rule-Breaking Behaviors (0.66), and lower correlations were found between Anxious/Depressed and Self-Regulation (0.14). **[*Table 5 near here*]**

Discussion

The primary purpose of this study was to develop the Nemad Electronic Mental-Health Assessment Devices for Children (NEMAD-C) to provide a reliable and valid Persian-language screening instrument for mental health problems among Iranian children that would be simple to administer and not require skills of a trained interviewer. Comparing expert panel form with the exploratory factor analysis and confirmatory factor analysis showed that in both PR and TR forms, expert panel form had better fit indexes. As such, expert panel form was used for further analysis. Internal consistency indices for all subscales in PR and TR, except self-regulation subscale in PR, were considered good and even excellent in some cases based on DeVellis (2012). Self-regulation subscale in PR had fairly low internal consistency. This low internal consistency may be attributable to low variance of item 10 ("He/she needs help with the shower") – 73% responded "0" to this item, indicating "never" performed – item 51 ("He/she needs help to dress up") – 74% responded "0" to this item, indicating "never" performed – and item 103 ("He/she has unstable friendships") – 61% responded "0" to this item, indicating "never" performed. The low variance of these items may account for the low internal consistency. These items may simultaneously represent several aspects of mental health problems which may be considered an issue regarding to content validity. Because of that, it is suggested that these items replace or re-word to enhance their clarity and psychometric properties in the revised version of the NEMAD-C.

The NEMAD-C-PR and TR possesses some preliminary evidence of convergent validity. As anticipated, positive relationships were found between the subscales of the NEMAD-C-PR and CBCL and also positive relationships were found between NEMAD-C-TR and TRF. Relationships of the subscales with CBCL and TRF suggest predictive validity for NEMAD-C-PR and TR. The weak significant correlations may have occurred because these constructions have weak correlations in real life situations. For example, in the present study, anxiety (ANX) was less strongly correlated with the rule-breaking behaviors (RB) (0.28). Many studies (e.g. Cooley et al., 2017) suggest that anxiety symptoms may attenuate the bidirectional relationships between physical and relational forms of aggression and peer victimization during middle childhood.

The present study has a few implications for the theory and practice of child psychology and mental health. In terms of theory, findings from this study expand the existing repertoire of behavioral problem indicators, while providing additional scales for self-harm, child abuse, and self-regulation and verify the latent structure underlying these constructs. This research offers a theoretical contribution to the conceptualization and operationalization of child outcome research in pediatric psychology; suggesting that assessing children problem behavior is a promising endeavor that warrants further attention in both research and practice. Regarding the

practice of child psychology and mental health, findings suggest that NEMAD-C is promising instrument for assessing problem behaviors, risk and psychopathology. Furthermore, using online questionnaires will deliver significant savings in human resources and reduction in time-to-hire.

The development of a universal assessment tool which is operable in the schools and resolve diverse range of behavioral problems could be a crucial segment in the development of an integrated approach for early identification of mental health conditions in children (Soneson et al., 2020; Humphrey & Wigelsworth, 2016). As schools have limited budget and financial resource for mental health services, universal screening systems can ease school staff's concerns toward additional material and human resources, costs and time (Soneson et al., 2020). The main logic is that universal screening should lead to early detection and primary tailored interventions for emergent mental health issues (Elliott et al., 2018; Eklund et al., 2009), which is more cost-effective than a clinically based approach for treatment of mental disorders and long-term behavioral difficulties. Therefore, it is suggested that universal screening becomes an inseparable part of school's services, and NEMAD-C which is a culturally adapted and validated screening instrument, could be used by professionals working in the area of mental health, by those in the primary health care setting.

Study limitations and future directions

Although our study provides valuable source of evidence on development of a culturally appropriate mental health screening tool for Iranian school-age children, it is not without limitations. First, demonstrable temporal stability (e.g., test/retest reliability) is lacking in this study which may affect our findings. According to the widely accepted view (Bland & Altman, 1986), high test re-test reliability implicates the internal validity of an instrument and assure that the measurements acquired in certain condition remain accurate and stable over time. Test-retest reliability assessment has a vital role in the development of psychometric instruments and screening tools, contributing to assure that measurement variation is due to replicable differences among individuals irrespective of contextual and demographic factors (Aldridge et al., 2017).

Second, many researchers in this field, particularly those who work with questionnaires and quantitative surveys, are probably aware of possibility of participant bias, social desirability and other biases on data quality. As a result, there may be an increased likelihood of some limitations which may affect the validity of the findings (Mahudin et al., 2012). In future investigations, we should address this issue by evaluating the influence of social context and social desirability. As such, these limitations could make necessary the use of social-desirability scales (Crowne & Marlowe, 1960), to control the social desirability response bias and other

types of biases that may limit the validity of a study (Petróczi & Nepusz, 2011). Future studies may also include other objective or independent measures to enrich the subjective evaluation of the variables studied in the development of the new instruments and to improve the interpretation and analysis of the results. Another limitation is that the number of the negatively-worded items and reverse-coded items (10 items in the parent form and 8 items in the teacher form) is very small compared to the positively-worded items. The imbalance between positively-worded and reverse-coded items would be redressed in the revised version of the NEMAD-C. Finally, web-based screening should be considered as a limitation in this study. Although time and cost savings are the most important advantage of this approach for data collection, its limitations should be carefully considered. Many researchers, for example, have mentioned that common concerns about web-based screening are coverage bias (i.e., bias due to sampled individuals not having the online platforms) and nonresponse bias (i.e., bias due to participants of a survey differing from those who did not respond about demographic) (Felderer et al., 2019). Alternatively, to minimize these issues for future investigations, researchers would benefit from in-person surveys or survey interviews in data collection methods. Although more time-consuming and costly, these methods alleviate problems related to confidentiality and the potential challenges for coverage and nonresponse bias (Reed et al., 2011). Therefore, where applicable, in-person assessments or interviews should be given priority in future directions for research in this field rather than electronic assessment systems.

References

Achenbach, T. M. (1991a). *Manual for the child behavior checklist/4-18 and 1991 profile*. University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 profile*. University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the child behavior checklist and revised behavior profile*. Burlington: University of Vermont, Department of Psychiatry.

Akbari Zardkhaneh, S., Poursharifi, H., Rajabi, M., & Yaghoubi, H. (2021). Mental Health of Iranian College Students in a Nationally Representative Sample of over 74,500 Individuals: Scale Development and Standardization. *Journal of College Student Psychotherapy*, *35*(2),1-20. https://doi.org/10.1080/87568225.2021.1911527

Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing test-retest reliability of psychological measures: Persistent methodological problems. *European Psychologist, 22*(4), 207-218.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Andrews, G., Sanderson, K., Corry, J., and Lapsley, H. M. (2000). Using epidemiological data to model efficiency in reducing the burden of depression. *The Journal of Mental Health Policy and Economics*, *3*, 175-186. https://doi.org/10.1002/mhp.96

Arslan, G. (2018). School-based social exclusion, affective wellbeing, and mental health problems in adolescents: A Study of mediator and moderator role of academic self-regulation. *Child Indicators Research*, *11*(3), 963-980. https://doi.org/10.1007/s12187-017-9486-3

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating casual models. *Journal of Marketing Research*, *19*(4), 425-430. https://doi.org/10.1177%2F002224378201900404

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588-606.

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307-310.

Brayden, R. M., Altemeier, W. A., Dietrich, M. S., Tucker, D. D., Christensen, M. J., McLaughlin, F. J., & Sherrod, K. B. (1993). A prospective study of secondary prevention of child maltreatment. *The Journal of Pediatrics*, *122*(4), 511-516. https://doi.org/10.1016/S0022-3476(05)83528-0

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504. https://doi.org/10.1080/10705510701301834

Cooley, J. L., Frazer, A. L., & Fite, P. J. (2017). Anxiety symptoms as a moderator of the reciprocal links between forms of aggression and peer victimization in middle childhood. *Aggressive Behavior*, *43*(5), 450-459.

Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology, *Journal of Consulting Psychology*, *24*(4), 349-354. https://doi.org/10.1037/h0047358

DeVellis, R.F. (2012). *Scale development: theory and applications*. Los Angeles: Sage.

Dowdy, E., Kamphaus, R. W., Twyford, J. M., Dever, B. V. (2014). *Handbook of School Mental Health*. Boston, MA: Springer.

Drummond, T. (1994). *The Student Risk Screening Scale (SRSS)*. Grants Pass, OR: Josephine County Mental Health Program.

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., et al. (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. *The California School Psychologist*, *14*(1), 89-95. https://doi.org/10.1007/BF03340954

Elliot, S. N., & Gresham, F. (2008). *Social Skills Improvement System Rating Scales*. Minneapolis, MN: NCS Pearson.

Elliott, S. N., Davies, M. D., Frey, J. R., Gresham, F., & Cooper, G. (2018). Development and initial validation of a social emotional learning assessment for universal screening. *Journal of Applied Developmental Psychology*, *55*, 39-51. https://doi.org/10.1016/j.appdev.2017.06.002

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.

Felderer, B., Kirchner, A., & Kreuter, F. (2019). The effect of survey mode on data quality: Disentangling nonresponse and measurement error bias. *Journal of Official Statistics*, *35*(1), 93-115. https://doi.org/10.2478/jos-2019-0005

Franklin, S. B., Gibson, D., Robertson, P., Pohlman, J. T., & Fralish, J. S. (1995). Parallel analysis: a method for determining significant principal components. *Journal of Vegetation Science*, *6*, 99-106.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*(2), 117-135. doi:https://doi.org/10.1016/j.jsp.2006.05.005

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

Haile, Y. G., Alemu, S. M., & Habtewold, T. D. (2017). Common mental disorder and its association with academic performance among Debre Berhan Universot students, Ethiopia. *International Journal of Mental Health Systems*, *11*(1), 1-11. https://doi.org/10.1186/s13033-017-0142-6

Hawton, K., Bergen, H., Cooper, J., Turnbull, P., Waters, K., Ness, J., et al. (2015). Suicide following self-harm: findings from the Multicentre Study of self-harm in England, 2000-2012. *Journal of Affective Disorders*, *175*, 147-51. https://doi.org/10.1016/j.jad.2014.12.062

Hu, L., & Bentler, P. M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology*, *108*(3), 663-676. https://doi.org/10.1016/j.jecp.2010.06.005

Humphrey, N., & Wigelsworth, M. (2016). Making the case for universal school-based mental health screening. *Emotional and Behavioural Difficulties*, *21*(1), 22-42. https://doi.org/10.1080/13632752.2015.1120051

Hunter, R. S., Kilstrom, N., Kraybill, E. N., & Loda, F. (1978). Antecedents of child abuse and neglect in premature infants: a prospective study in a newborn intensive care unit. *Pediatrics*, *61*(4), 629-635.

Kaiser, B. N., Ticao, C., Anoje, C., Minto, J., Boglosa, J., & Kohrt, B. A. (2019). Adapting culturally appropriate mental health screening tools for use among conflict-affected and other vulnerable adolescents in Nigeria. *Global Mental Health*, *6*, e10. http://doi.org/10.1017/gmh.2019.8

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2: Behavioral and Emotional Screening System*. Minneapolis, MN: Pearson.

Kettler, R. J., Glover, T. A., Albers, C. A., & Feeney-Kettler, K. A. (Eds.). (2014). *School psychology book series. Universal screening in educational settings: Evidence-based decision making for schools.* Washington, DC: American Psychological Association. https://psycnet.apa.org/doi/10.1037/14316-000

Kirmayer, L. J., Gomez-Carrillo, A., & Veissière, S. (2017). Culture and depression in global mental health: An ecosocial approach to the phenomenology of psychiatric disorders. *Social Science & Medicine*, *183*, 163–168. https://doi.org/10.1016/j.socscimed.2017.04.034

Klevens, J., & Whitaker, D. J. (2007). Primary prevention of child physical abuse and neglect: gaps and promising directions. *Child Maltreatment*, *12*(4), 364-377. https://doi.org/10.1177%2F1077559507305995

Loewenthal, K. M. (2019). *Intersections among religion, culture, gender, and mental health.* In J. A. Mena & K. Quina (Eds.), *Integrating multiculturalism and intersectionality into the psychology curriculum: Strategies for instructors* (p. 143–156). American Psychological Association. https://doi.org/10.1037/0000137-012

Maalouf, F. T., Alamiri, B., Atweh, S., Becker, A. E., Cheour, M., Darwish, H., ... & Khoury, B. (2019). Mental health research in the Arab region: challenges and call for action. *The Lancet Psychiatry*, *6*(11), 961-966. https://doi.org/10.1016/S2215-0366(19)30124-5

Mahmood, S. I., Abdul-Daim, S., Peninga, M. H., & Adema, J. S. (2015). The transferability of western concepts to other cultures: Validation of the Zuckerman-Kuhlman personality questionnaire in a Saudi Arabic context. *Medical Teacher*, *37*, S67–S74. https://doi.org/10.3109/0142159X.2015.1006606

Mahudin, N. D. M., Cox, T., & Griffiths, A. (2012). Measuring rail passenger crowding: scale development and psychometric properties. *Transportation Research Part*, *15*, 38–51. https://doi.org/10.1016/j.trf.2011.11.006

McClelland, M.M., Ponitz, C.C., Messersmith, E.E. and Tominey, S. (2010). Self-Regulation. *In The Handbook of Life-Span Development* (eds R.M. Lerner, M.E. Lamb and A.M. Freund). https://doi.org/10.1002/9780470880166.hlsd001015

McCoy, D. C. (2019). Measuring young children's executive function and self-regulation in classrooms and other real-world settings. *Clinical Child and Family Psychology Review*, *22*(1), 63-74. https://doi.org/10.1007/s10567-019-00285-1

Miller, D. N. (2021). *Child and adolescent suicidal behavior: Schoolbased prevention, assessment, and intervention*. New York: The Guilford Press.

Milner, J. S., Gold, R. G., Ayoub, C., & Jacewitz, M. M. (1984). Predictive validity of the Child Abuse Potential Inventory. *Journal of Consulting and Clinical Psychology, 52*(5), 879-884. https://psycnet.apa.org/doi/10.1037/0022-006X.52.5.879

Muir, R. C., Monaghan, S. M., Gilmore, R. J., Clarkson, J. E., Crooks, T. J., & Egan, T. G. (1989). Predicting Child Abuse and Neglect in New Zealand. *Australian and New Zealand Journal of Psychiatry*, *23*(2), 255-260. https://doi.org/10.3109%2F00048678909062143

Newbury, J. B., Arseneault, L., Moffitt, T. E., Caspi, A., Danese, A., Baldwin, J. R., & Fisher, H. L. (2018). Measuring childhood maltreatment to predict early-adult psychopathology: Comparison of prospective informant-reports and retrospective self-reports. *Journal of Psychiatric Research*, *96*, 57-64. https://doi.org/10.1016/j.jpsychires.2017.09.020

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

Palacios-Barrios, E. E., & Hanson, J. L. (2019). Poverty and self-regulation: Connecting psychosocial processes, neurobiology, and the risk for psychopathology. *Comprehensive Psychiatry*, *90*, 52-64. https://doi.org/10.1016/j.comppsych.2018.12.012

Patel, V. (2001). Cultural factors and international epidemiology. *British Medical Bulletin*, *57*, 33– 45. https://doi.org/10.1093/bmb/57.1.33

Patel, V., Flisher, A., Nikapota, A., & Malhotra, S. (2008). Promoting child and adolescents mental health in low and middle income countries. *Journal of Child Psychology and Psychiatry*, *49*, 313–334. https://doi.org/10.1111/j.1469-7610.2007.01824.x

Patton, G. C., Coffey, C., Sawyer, S. M., Viner, R. M., Haller, D. M., Bose, K., et al. (2009). Global patterns of mortality in young people: a systematic analysis of population health data. *Lancet*, *374*, 881-92. https://doi.org/10.1016/S0140-6736(09)60741-8

Peters, R., & Barlow, J. (2003). Systematic review of instruments designed to predict child maltreatment during the antenatal and postnatal periods. *Child Abuse Review*, *12*(6), 416-439. https://doi.org/10.1002/car.821

Petróczi, A., & Nepusz, T. (2011). Methodological considerations regarding response bias effect in substance use research: is correlation between the measured variables sufficient?. *Substance Abuse Treatment, Prevention, and Policy*, *6*(1), 1-11. https://doi.org/10.1186/1747-597X-6-1

Reed, L. L., Vidaver-Cohen, D., & Colwell, S. R. (2011). A new scale to measure executive servant leadership: development, analysis, and implications for research. *Journal of Business Ethics*, *101*, 415–434. http://dx.doi.org/10.1007/ s10551-010-0729-1.

Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology*, *45*(2), 193-223. https://doi.org/10.1016/j.jsp.2006.11.003

Shahrivar, Z., Tehrani-Doost, M., Khorrami Banaraki, A., Mohammadzadeh, A. and Happe, F. (2017). Normative data and psychometric properties of a Farsi translation of the strange stories test. *Autism Research*, *10*, 1960-1967. https://doi.org/10.1002/aur.1844

Soneson, E., Howarth, E., Ford, T., Humphrey, A., Jones, P. B., Coon, J. T., ... & Anderson, J. K. (2020). Feasibility of school-based identification of children and adolescents experiencing, or at-risk of developing, mental health difficulties: a systematic review. *Prevention Science*, 1-23. https://doi.org/10.1007/s11121-020-01095-6

Stephens-Reicher, J., Metcalf, A., Blanchard, M., Mangan, C., & Burns, J. (2011). Reaching the hard to reach: how information technologies can reach young people at greater risk of mental health difficulties. *Australian Psychiatry*, *19*(1), 58-61. https://doi.org/10.3109/10398562.2011.583077

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structure equation models. In K. A. Bollen & J.S. Long (Eds.). *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Toosi, F., Rahimi, C., & Sajjadi, S. (2017). Psychometric properties of Beck Depression Inventory-II for high school children in Shiraz City, Iran. *International Journal of School Health*, *4*(3). https://dx.doi.org/10.5812/intjsh.41069

Trautmann, S., Rehm, J., & Wittchen, H. U. (2016). The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders?. *EMBO Reports*, *17*(9), 1245-1249. https://doi.org/10.15252/embr.201642951

Verlenden, J., Naser, S., & Brown, J. (2020). Steps in the Implementation of Universal Screening for Behavioral and Emotional Risk to Support Multi-Tiered Systems of Support: Two Case Studies. *Journal of Applied School Psychology*, 1-39. https://doi.org/10.1080/15377903.2020.1780660

Walker, H. M., Nishioka, V. M., Zeller, R., Severson, H. H., & Feil, E. G. (2000). Causal factors and potential solutions for the persistent underidentification of students having emotional or behavioral disorders in the context of schooling. *Assessment for Effective Intervention*, *26*(1), 29-39. https://doi.org/10.1177%2F073724770002600105

Walker, H., & Severson, H. (1992). *Systematic screening for behavior disorders: SSBD*. Longmont, CO: Sopris West.

Whiteford, H. A., Ferrari, A. J., Degenhardt, L., Feigin, V., & Vos, T. (2015). The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PloS One*, *10*(2), e0116820. https://doi.org/10.1371/journal.pone.0116820

Woodward, L. J., Lu, Z., Morris, A. R., & Healey, D. M. (2017). Preschool self regulation predicts later mental health and educational achievement in very preterm and typically developing children. *The Clinical Neuropsychologist*, *31*(2), 404-422. https://doi.org/10.1080/13854046.2016.1251614

World Health Organization. (2006). WHO-AIMS report on mental health system in the Islamic Republic of Iran. *Tehrran, Iran: World Health Organization—Assessment Instrument for Mental Health Systems*.

Tables:

Table 1. Summary of Item analysis results for PR and TR form

Scale	Item	IM	SK	KU	SD	SMC	CITC
PR	Has difficulties in following orders	0.80	0.99	-0.03	0.92	0.43	0.64
	Worries about reading, writing, or speaking in crowd	0.66	1.30	0.70	0.90	0.39	0.58
	Gets angry easily	0.91	0.91	-0.33	1.04	0.49	0.66
	Acts slowly	0.82	0.99	-0.12	0.97	0.41	0.56
	Gets tired so easily during homework	0.96	0.74	-0.58	0.98	0.59	0.73
	Shows signs of skin picking	0.18	3.26	10.86	0.56	0.35	0.54
TR	Has difficulties in concentrating on homework	1.01	0.70	-0.73	1.05	0.46	0.60
	Refuses to attend school due to emotional distress	0.34	2.53	6.27	0.71	0.56	0.69
	Uses shouting as a way to reach his/her wishes	0.51	1.62	1.84	0.84	0.60	0.74
	Has lost interest in his/her favorite activities	0.47	1.76	2.67	0.75	0.62	0.76
	Doing his/her homework carelessly	1.02	0.74	-0.64	1.04	0.75	0.85
	Injures parts of his/her body	0.20	3.77	14.66	0.60	0.63	0.77

Item Mean (IM), Skewness (SK), Kurtosis (KU), Standard Deviation (SD), Squared Multiple Correlation (SMC), Corrected Item Total Correlation (CITC)

Table 2. Model of fit statistics for PR

	Model	CFI	NNFI	RMSEA	IFI	RMSEA (90% CI)	SRMR
PR	EFA	0.97	0.94	0.054	0.97	0.054-0.055	0.069
	EP	0.96	0.96	0.059	0.96	0.065-0.066	0.063
TR	EFA	0.98	0.98	0.071	0.98	0.070-0.071	0.074
	EP	0.98	0.98	0.064	0.98	0.098-0.099	0.084

IFI: Incremental Fit Index, CFI: Comparative Fit Index, NNFI: Non-Normed Fit Index, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual, EFA: Exploratory Factor Analysis, EP: Expert Panel

Table 3. Descriptive properties, internal consistency, and inter-correlations of the PR

	ADHD	DB	ANX	DEP	SH	AAD	SR	CAR
M	8.14	6.16	6.38	6	2.19	10.85	13.36	3.63
SD	7.28	6.54	5.82	5.59	3.81	7.16	5.59	3.95
CA	0.89	0.89	0.83	0.82	0.81	0.84	0.65	0.73
Item range	0.46-0.84	0.19-0.91	0.27-0.99	0.26-0.88	0.10-0.31	0.50-1.67	0.43-1.81	0.09-0.72
ICR range	0.50-0.66	0.35-0.69	0.30-0.59	0.34-0.62	0.34-0.54	-0.10-0.73	0.14-0.45	0.30-0.48
Correlation								
ADHD (1)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
DB (2)	0.81							
ANX (3)	0.70	0.69						
DEP (4)	0.75	0.72	0.76					
SH (5)	0.59	0.68	0.60	0.61				
AAD (6)	0.78	0.67	0.61	0.73	0.44			
SR (7)	0.48	0.45	0.49	0.42	0.28	0.55		
CAR (8)	0.68	0.68	0.70	0.72	0.69	0.58	0.39	-

M: Mean, SD: Standard Deviation, IR: Item Range, CA: Cronbach's Alpha, ICR: Item Correlation Range, CAR: Child Abuse Risk, SH: Self- Harm, ANX: Anxiety, DB: Disruptive Behavior disorders, ADHD: Attention deficit/hyperactivity disorder, DEP: Depression, AAD: Academic Achievement Deficit, SR: Self-Regulation.

All the correlations were significant at $p < 0.05$

Table 4. Descriptive properties, internal consistency, and inter-correlations of the TR

	ADHD	DB	ANX	DEP	SH	AAD	SR	CAR
M	6.92	5.99	5.24	5.67	1.56	9.36	8.44	3.57
SD	7.04	6.36	5.80	6.10	3.62	8.60	5.22	5.14
CA	0.92	0.91	0.90	0.91	0.93	0.96	0.80	0.90

Item range	0.52-1.01	0.40-0.90	0.27-0.84	0.33-0.99	0.19-0.28	0.79-1.06	0.50-1.25	0.17-0.61
ICR range	0.60-0.79	0.28-0.81	0.58-0.74	0.60-0.76	0.72-0.81	0.78-0.87	0.39-0.56	0.57-0.75
Correlation								
ADHD (1)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
DB (2)	0.89							
ANX (3)	0.76	0.76						
DEP (4)	0.77	0.74	0.89					
SH (5)	0.65	0.74	0.77	0.70				
AAD (6)	0.80	0.67	0.67	0.79	0.44			
SR (7)	0.72	0.60	0.63	0.72	0.42	0.80		
CAR (8)	0.75	0.79	0.84	0.84	0.87	0.61	0.57	-

M: Mean, SD: Standard Deviation, CA: Cronbach's Alpha, IR: Item Range, ICR: Item Correlation Range, CAR: Child Abuse Risk, SH: Self-Harm, ANX: Anxiety, DB: Disruptive Behavior disorders, ADHD: Attention deficit/hyperactivity disorder, DEP: Depression, AAD: Academic Achievement Deficit, SR: Self-Regulation.

All the correlations were significant at $p < 0.05$

Table 5. Correlations between NEMAD-C and CBCL

NEMAD-C		**CBCL/TRF**							
		AD	WD	SC	SP	TP	AP	RB	AB
PR	ADHD	0.38**	0.34**	0.31**	0.50**	0.44**	0.63**	0.51**	0.61**
	DB	0.49**	0.41**	0.31**	0.52**	0.53**	0.55**	0.59**	0.74**
	ANX	0.60**	0.53**	0.37**	0.51**	0.46**	0.43**	0.28**	0.41**
	DEP	0.61**	0.56**	0.48**	0.60**	0.52**	0.61**	0.42**	0.58**
	SH	0.37**	0.33**	0.26**	0.42**	0.48**	0.43**	0.48**	0.58**
	AAD	0.34**	0.33**	0.25**	0.45**	0.34**	0.61**	0.38**	0.48**
	SR	0.35**	0.28**	0.23**	0.37**	0.29**	0.43**	0.29**	0.38**
	CAR	0.40**	0.28**	0.23**	0.37**	0.35**	0.25**	0.21**	0.32**

TR	ADHD	0.31**	0.25**	0.33**	0.59**	0.53**	0.53**	0.65**	0.68**
	DB	0.24**	0.14**	0.24**	0.62**	0.45**	0.46**	0.66**	0.73**
	ANX	0.44**	0.35**	0.30**	0.49**	0.31**	0.33**	0.32**	0.18**
	DEP	0.39**	0.42**	0.31**	0.52**	0.38**	0.49**	0.40**	0.22**
	SH	0.30**	0.22**	0.26**	0.48**	0.40**	0.27**	0.46**	0.41**
	AAD	0.22**	0.31**	0.28**	0.40**	0.42**	0.55**	0.42**	0.27**
	SR	0.14**	0.27**	0.21**	0.40**	0.33**	0.51**	0.40**	0.26**
	CAR	0.34**	0.34**	0.33**	0.57**	0.47**	0.41**	0.54**	0.41**

CAR: Child Abuse Risk, SH: Self- Harm, ANX: Anxiety, DB: Disruptive Behavior disorders, ADHD: Attention deficit/hyperactivity disorder, DEP: Depression, AAD: Academic Achievement Deficit, SR: Self-Regulation, AD: Anxious/Depressed, WD: Withdrawn/Depressed, SC: Somatic Complaints, SP: Social Problems, TP: Thought Problems, AP: Attention Problems, RB: Rule-Breaking Behavior, AB: Aggressive Behavior, CBCL: Child Behavior Checklist, TRF: Teacher Form.

** All the correlations were significant at $p < 0.01$