BG Research Online

This version may differ slightly from the final published version.

For enquiries about BG Research Online email bgro@bishopg.ac.uk.

Assessing children at risk: Development and psychometric properties of the Motor Behavior Checklist

Abstract

Objective: Development and investigation of the psychometric properties of the Motor Behavior Checklist (MBC; Efstratopoulou, Jansen, & Simons, 2012), as a new screening instrument for assessing children's deviant behaviors.

Design: Confirmatory Factor Analyses (CFAs) were used to examine the construct validity of the MBC.

Method: Instrument development included physical educators' reports, official psychiatric diagnoses and a sorting task by a team of experts. A normative database of primary students (*N*=841), who were rated by their physical educators in school settings, were used to identify the factor structure of the list and to investigate the internal consistency, the reproducibility and the inter-rater agreement.

Results: A series of Confirmatory Factor Analyses (CFAs) revealed a second order model with two (Externalizing and Internalizing) broadband domains and seven problems scales: Rules breaking (7 items), Low energy (4 items), Stereotyped behaviors (2 items), Hyperactivity/Impulsivity (14 items), Lack of Attention (10 items), Lack of Social interaction (10 items), and Lack of Self regulation (12 items).

Conclusions: Psychometric results supported the model suggesting that MBC for children is a new promising instrument homogenous in content, with high temporal stability and significant correlation agreement that can provide useful and reliable ratings on behavioral and emotional problems in children when used by physical educators in school settings.

Keywords: motor related behaviors, checklist, physical education, cluster analysis, confirmatory factor analysis, reliability

Assessing children at risk: Development and psychometric properties of the Motor Behavior Checklist

Introduction

The importance of early diagnosis

Education research indicates that early identification for emotional and/or behavioral problems can help to minimize the long-term harm of mental health disorders and reduce the overall healthcare burden and costs (Aos, Lieb, Mayfield, Miller, & Pennucci, 2004). Detection efforts are particularly critical during the early educational years, when students are most amenable to change in behavioral, social, and academic arenas and before students at risk for emotional and behavioral disorders (EBD) and children with autistic spectrum disorder (ASD), experience negative outcomes within and beyond the school setting (Landrum, Tankersley, & Kauffman, 2003; Lane, 2003; Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005; Volkmar, Lord, Bailey, Schultz, & Klin, 2004).

Students with EBD who do not receive necessary support often experience a host of negative outcomes, including peer and teacher rejection, academic underachievement, school dropout, substance abuse, depression, unemployment, and involvement in the juvenile justice system (Bullis & Yovanoff, 2006; Morris, Shah, & Morris, 2002; Wagner & Davis, 2006; Wagner, Newman, Cameto, Levine, & Garza, 2006; Zigmond, 2006). Given the costs associated with EBD, to students themselves, their families, and society as a whole, it is not surprising that reducing the incidence of EBD through systematic screening and comprehensive intervention efforts is a growing area of interest to educational research (Kauffman & Landrum, 2009; Lane, 2007; Nelson, Babyak, Gonzalez, & Benner, 2003).

Existing instruments and agreement between rating sources

Short measures with known reliability and validity are crucial for triggering the classroom practices and school services necessary to ensure that all children succeed in school (DiStefano & Kamphaus, 2007). A range of formal and informal measures and techniques, especially rating scales and behavior checklists, are used to gather adults' perceptions of children's behaviors at home and at school. The main disadvantage of the existing instruments is that they are focusing mainly on specific disorders like the Social Skills Rating System–Student Form (SSRS-SF; Gresham & Elliott, 1990), and the Social Anxiety Scales for children and adolescents (SASC-R, La Greca, 1999), or the Conners' Rating Scales (2000) and the ADHD Rating Scale (DuPaul,

1998), when there is a great need for a more global and comprehensive assessment concerning deviant behaviors of typical school-age population.

Considering this need, historically, assessment systems such as the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach, 1991) and individual instruments such as the Revised Behavior Problem Checklist (Quay & Peterson, 1987) or the Behavioral Assessment System for Children (Reynolds & Kamphaus, 2004), have focused on individual deficits, problems, or pathologies when recent instruments have begun targeting behavioral and emotional strengths and characteristics, such as the Behavioral and Emotional Rating Scale–Second Edition (BERS-2; Epstein, 2004).

Despite the usefulness of rating techniques for describing children's behaviors, the relatively modest agreements among rating sources (Verhulst & Akkerhuis, 1989) raise questions about the validity of information and the importance of context or setting effects on children's behavior. The frequency, base rate, and conspicuousness of behaviours may affect the degree of concordance among informants (Kolko & Kazdin, 1993). Considerable literature addresses issues of method effects in cross-informant studies, and there are many explanations for rater disagreement (e.g., Gadow, Drabick, Loney, Sprafkin, Salisbury, Azizian, et al., 2004; Drabick, Gadow, & Loney, 2008), but important to the understanding of the etiology of syndromes is the notion that different sources of information may differentially identify specific behaviors with different etiologies or provide unique information about the psychopathogenesis of seemingly similar clinical phenotypes (e.g., Boomsma, van Beijsterveldt, & Hudziak, 2005; Hudziak, Derks, Althoff, Copeland, & Boomsma, 2005; Martin, Scourfield, & McGuffin, 2002). In general, concordance has been found to be higher when informants have similar relationships with the children being rated (e.g., teachers, educators) than between teachers and parents which represent different roles (Achenbach, McConaughy, & Howell, 1987; Greenbaum, Dedrick, Prange, & Friedman, 1994). Thus, there is stronger agreement between parents than between parents and teachers, suggesting that there may be differences in raters' frames of reference and/or that children's behaviors vary in different settings.

The role of physical educators in the assessing procedure

Although classroom teachers are in an excellent position to provide information about the child's behavior, they observe their children mainly during lesson at class settings. In contrast with class settings, Physical Education (PE) lessons and group play situations provide a unique

opportunity to observe a child interacting with his/her peers, co-operating or just being on his/her own. The fact that physical educators spend a lot of time with the children and have the flexibility to work with them and observe their behaviors in several ways (e.g., structured lessons or free play situations) and several different settings (inside or outside the classroom, at the playground or at the school-yard), give them the opportunity to distinguish between maladaptive and general age-related motor behaviors. Evidence for the presence of externalizing and/or internalizing symptoms can be obtained in multiple active situations, and a number of behavioral symptoms can be observed during PE classes and team games (Kashani, Allan, Beck, Bledsoe, & Reid, 1997). In addition, results on observational studies suggest that explicit behavioral symptoms can be systematically observed during standardised play procedures (Mol Lous, Wit, De Bruyn, & Riksen-Walrawen, 2002) and educators who observe different aspects of children during their lessons are able to identify young children at high risk for school adjustment problems related to attention, conduct, learning, and mood with a great deal of accuracy (Flanagan, Bierman, & Kam, 2003).

Physical education teachers have the knowledge and the skills to focus on the "warning sings" of atypical motor behaviors providing useful information about the development of school-aged children. However, there are only a few instruments that use the physical educators as main source of information about children's development and the majority of them are focusing on movement and motor coordination problems like the Bruininks-Oseretsky Test of Motor Proficiency (BOT-2; Bruininks & Bruininks, 2005), the Test of Motor Development (TGMD; Ulrich, 2000), or the Movement Assessment Battery for Children (MABC-II; Henderson, & Sugden, 2007), which assess gross and fine motor skills, balance, levels of motor skill development as part of psychological test batteries, for making decisions about educational placement, developing and evaluating intervention programs.

In addition, none of the existing instruments for physical educators assess a wide array of children's problematic behaviors, as most of them are focusing only on specific disorders which are highly connected with performance in sports or with class management in school settings (STAIC; Spielberger & Edwards, 1973; PECI; Kullina, 2003).

The current study

Considering the need for instruments that are practical for wide-scale school use by physical educators, assess a wide array of children's behaviors and possess supportive psychometric evidence, the current study aimed to develop a practical and reliable instrument that addresses these

points. The instrument was designed to be used by physical educators in school settings for identifying children at-risk for behavioral and emotional problems on the basis of their motor related behaviors and refer them accurately for further assessment.

The research was conducted in two studies containing several separate substudies each. Study one aimed to develop the preliminary version of the Motor Behavior Checklist for children, and study two aimed to examine the psychometric properties of the new instrument, namely: the structure validity, the internal consistency, the test-retest and the inter-rater reliability.

Study one: Development of the MBC for children

Aim of study one

Aim of study one was the development of the preliminary list of MBC and it was conducted in two steps. Step one included the selection of Physical Educators reports about children's problematic behaviors in school settings, and step two included the selection of observable motor related behaviors from the official psychiatric criteria.

The Physical Educators' list

Sixty physical educators, 32 males (52%) and 28 females (48%), with mean age of 42.1 years (SD=6.3 years) and mean teaching experience 10.2 years (SD=3.4 years) in typical school populations at public elementary schools, were asked to describe the full range of children's deviant behaviors they observe among their pupils using a brief open-ended questionnaire. Physical educator's descriptions were screened by three experts in adapted physical activity and a list containing motor related behaviors unique to physical education class settings, was derived.

Selection of Official Diagnostic Criteria

Diagnostic criteria from the two most used official manuals in children's and adult psychopathology: Diagnostic and Statistic Manual, DSM–IV, (American Psychiatric Association, 2000) and the International Diagnostic Classification, IDC-10, (World Health Organization, 1992) were screened by three experts in order to select criteria that refer to motor related behaviors easy to be observed within a school environment. Criteria coming from different diagnoses but describing the same motor-related behaviors, due to high rates of symptom overlap (Klassen, Miller, & Fine, 2004), were used only once in the final list.

Grouping the items into categories

Diagnostic criteria combined with the physical educators' initial descriptions of deviant behavior were entered in a sorting task (Rosenberg & Jones, 1972; Rosenberg & Kim, 1975) in

which a sample of 50 physical education experts participated in order to investigate their perceptions of children's deviant behaviors and how these different forms of deviant behaviors are perceived as parts of a specific category. The sample consisted of 50 physical education teachers, 29 males (58%) and 21 females (42%) with mean age 38.4 years (*SD*=5.2 years) from which 42 (84%) had a master degree in education and 8 (16%) had a PhD in education. Participants asked to sort the items of the total list which was containing a combination of the PE reports and the diagnostic criteria, into different categories based on their perceived similarity in content with no limitations as to the number of categories or the number of items within each category. A similarity matrix of the sorted items was derived by counting the number of times participants sorted a pair of items in the same group. The similarity matrix was submitted to a hierarchical cluster analysis (using SPSS 15.0, 2006) in order to delineate subsets of similar items (called clusters) and the hierarchical structure among these clusters.

Results of study one

Based on Physical educator's reports, 65 different behaviors were formulated into items describing deviant children's' behaviors. A content analysis of the items revealed that 44 (68%) of the total 65 items, were statements about disobedience, negative reactions to rules and aggressive behaviors, 13 items (20%) were statements describing lack of concentration, attention deficits, careless mistakes, and impulsive motor-related behaviors and only eight statements (12%) referred to problematic social behaviors.

The screening of the Diagnostic manuals led to a list of 187 diagnostic criteria of which the majority (43%) referred to impulse-control disorders, namely oppositional defiant disorder (ODD), conduct disorder (CD), and attention deficit/hyperactivity disorder (ADHD), 21% came from pervasive developmental disorders, and 20% accounted for anxiety disorders.

Results of the complete linkage solution of the derived similarity matrix on the sum of the items (n=252) revealed two main clusters containing two and seven subcluster each. The similarity among the items in the nine subclusters ranged from 75% to 90%, implying that between 36 and 42 out of 50 physical educators sorted these items together in the same group. Hence, these items can be assumed to be homogeneous in content. At the second level of clustering, agreement among raters ranged between 54% and 71%. The seven subclusters that were formulated were: subcluster a (disobedience), subcluster b (aggression), subcluster c (hyperactivity), subcluster d (impulsivity), subcluster e (stereotyped), subcluster f (low energy),

subcluster g (inattention), subcluster h (isolation) and subcluster i (anxiety). During instrument development, every attempt was made to be sensitive to the varied contexts and children who participate in physical education classes and some of the selected items represented behaviors unique to physical education settings. <mark>Based on specific selection criteria the most representative items from each cluster were selected and a preliminary list was developed. The selection criteria included: a) priority to the items coming from physical educator's reports, b) items had to describe with clarity motor behaviors, and c) these motor behaviors had to be easily observe in school settings</mark>.

In this way, a preliminary 85-items list was developed and a Likert-like response scale format was used ranging from 0 (never) to 4 (almost always) for each of the items in order to give the rater the possibility to provide information about the frequency of the exhibited behavior.

Summary of study one

Analyzing the results from study one it was obvious that Physical Education teachers were able to observe deviant or problematic motor related behaviors among their students, and also to discern patterns of similar behavior among the items, leading to a meaningful hierarchical cluster analysis resembled the well-known distinction between externalizing and internalizing problem behavior. Based on the results from the cluster analysis and giving priority to the items that derived from physical educator's reports, a team of experts in adapted physical activity, selected items from each subcluster. A preliminary version of the Motor Behavior Checklist (MBC) for children containing 85 items was developed.

Study two: Psychometric properties of the MBC for children

Aim of study two

After instrument development, the aim of study two was to investigate the key psychometric properties, of the Motor Behavior Checklist (MBC) for children, namely: the structure validity, the internal consistency, the reproducibility and the inter-rater agreement, using a normative data of elementary school-aged children.

Participants

The data analysed was collected from a randomly selected sample ($N = 841$) of elementary school-aged children. School review broad approval was obtained as well as appropriate consent/assent from participants and their parents. The data collected was anonymous and only codes about demographic characteristics of the participants were used. The overall sample was

consisted of 421 (50, 1%) girls and 420 (49, 9%) boys, ranging in age from 6 to 11 years (M=8.4 years, SD=1.7 years) and they had the Greek nationality (99 %). The data derived from 35 typical Greek elementary schools widely spread across Greece selected so that the sample distribution would be representative of the urban and rural population. The schools were located in urban areas (63.3%) and in rural areas and islands (36,7%). The physical education teachers (N=62) of the schools who participated in this study were 35 females (56%) and 27 males (44%), with mean age 39.4 years (SD=6.2 years) and mean teaching experience 7.2 years (SD=3.4 years). The participant's physical educators of the schools were asked to randomly select four children (2 boys and 2 girls) from each class and rate them using the 85-items preliminary version of MBC for children. The initial data were randomly divided into two sample groups. Sample one (N=426) was used for examining the structure validity of the list and sample two (N=415) was used to assess the model fit. An overview of children's characteristics from confirmatory factor analyses and reliability studies are presented in table 1.

Table 1.

Method

The aims of study two were twofold: a) the investigation of the factorial structure via confirmatory analyses and based on these results select the best items, and b) the investigation of the reliability of the list. The investigation of the structure validity was conducted into three different methodological steps: First step was the initial examination of the factor structure of the list based on the cluster solution from study one. Second step was the selection of the items based on specific selection criteria, and third step was the investigation of the adequacy of the model fit. Concerning the reliability of the list, the internal consistency, the reproducibility and the inter-rater agreement were also examined.

Factor structure

The initial examination of the factor structure was based on the results of the cluster solution from study one, which reveled two main clusters containing 9 subclusters. Data from sample one (N=426) were used in order to ran a Confirmatory Factor Analysis (CFA) using Maximum-likelihood method (LISREL 8; Jöreskog & Sörbom, 1993). Modifications to the hypothesized factor model were made based on the correlations among factors.

Selection of the items

Taking under consideration the selection criteria proposed by Marsh, Ellis, Parada, Richards and Heubeck, (2005), the standardized loadings, the modifications indices and the item total correlations were examined in order to produce a more concise instrument within a parsimonious model. A new confirmatory analysis was performed in order to examine a second order factor model with seven factors using the 85 items of the list. In case of inadequate fit for the hypothesized model, modifications to the CFA were performed based on the analyses of items. The modifications made were based on high factor loadings, correlated uniqueness within each factor and inter correlations between items within each factor. More specifically, items that: a) best measured each factor having high standardized factor loadings (≥.50) and b) had minimal cross loadings on other factors as assessed via modification indices were chosen. Caution was taken not to reduce the number of subscales or the number of items within each scale so drastically that it led to construct under presentation that could mask the intended purpose and validity of the measure (Messick, 1995). The modifications made consisted of freeing up both crossloadings and correlated uniqueness within each factor until a reasonable fit was obtained.

Assessing Model Fit

To confirm the adequacy of the reduced model which revealed via the CFAs, data from sample two (*N*=415) were used to ran a second order factor confirmatory analysis using Maximum-likelihood (LISREL 8; Jöreskog & Sörbom, 1993). Because the chi-square statistic frequently yields false positives when large samples are analyzed (Floyd & Widaman, 1995), the Comparative fit index (CFI), the Normed fit index (NFI), the Goodness of Fit Index (GFI), the Root Mean of Approximation (RMSEA) and the Standardized Root mean Square Residual (SRMS) indexes were used to evaluate the fit of the data.

Internal consistency

Internal consistency was estimated in order to measure the extent to which items in a subscale are correlated (homogeneous) thus measuring the same concept. Cronbach's alpha coefficients were calculated of the subscales using the initial data of children (*N*=841). Alpha values estimated separately for the externalizing and internalizing scales and the 7 problem subscales.

Reproducibility

To determine test-retest reliability, Intraclass Correlation Coefficients (ICC) were calculated using a sample of 129 elementary school children, 61 girls (47%) and 68 boys (52%)

who were rated twice by their physical educators at schools. The children had mean age 8.51 years (*SD*=1. 75), 111 (86%) had the Greek nationality and they were attending nine typical elementary schools from Athens and Thessaloniki. The participant physical educators, seven females and four males, record their student's motor related behaviors during physical education lessons at school environment using the MBC for children twice within one month.

Inter-rater reliability

Data from twenty-two physical educators (14 females and 8 males) from 11 elementary schools from Athens and Thessaloniki, who rated 126 of their students using the MBC, was used in order to assess inter-rater reliability of the list. The participant children were 67 boys (53%) and 59 girls (47%), with mean age 8.51 years (*SD*=1.75) and 118 (92%) had the Greek nationality. From each school, two physical educators working independently with the same students but in different settings where involved. One was the traditional physical educator teacher working only in movement situations, involving sports, and the second one was working in both settings, inside the class, giving theoretical information about regulations in sports and nutrition, and outside the class working in team games.

Results of study two

Factor structure

Analyzing the results from the nine factors model, derived from the results of the cluster analysis in study one, the estimated correlation matrixes among the factors revealed high correlation between factor Disobedience and factor Aggressiveness (r=.98) and between factor Hyperactivity and factor Impulsivity (r=.95). Due to high correlations, it was assumed that these factors were so similar in content describing and assessing aspects of the same construct and therefore had to be reduced into two. Factor Disobedience (8 items) and factor Aggressiveness (7 items) were reduced into one factor containing the sum of their items (n=15 items) and it was named: Rules Breaking as the most of the items were describing aggressive behaviors mainly connected with disobedience and violation of rules in school environment. Factor Hyperactivity (6 items) and factor Impulsivity (9 items) were also reduced into one factor containing the sum of their items (n=15 items) and named: Hyperactivity/Impulsivity containing items describing hyperactive and impulsive behaviors. The other five factors were: Low energy (n=4 items), with items describing decreased activity, Stereotyped behaviors (n=6 items), with items describing repeated patterns of activity, Lack of attention (n=10 items), containing items describing problems in attention and lack of concentration,

Lack of social interaction (n=16 items), containing items describing problems in communication and social interaction with teachers and peers, and Lack of self regulation (n=19 items), containing mainly items describing anxiety and inability of the child to regulate behavior.

Selection of the items

A new second order CFA indicated the existence of two higher orders Externalizing and Internalizing factors containing the seven problems subscales. After deleting the items with low loadings and cross loadings, a 59-items list was developed. More specifically, for the reduced model, the factor Rules breaking was consisted of 7 items, the factor Hyperactivity/impulsivity was consisted of 14 items, the factor Lack of attention was consisted of 10 items, the factor Low energy was consisted of 4 items, the factor Stereotyped behavior was consisted of 2 items, the factor Lack of social interaction was consisted of 10 items and the factor Lack of self regulation was consisted of 12 items. The reduced model presented an optimal level of fit to the data (RMSEA=.074, CFI=.97, NFI=.96, GFI=.93, and SRMS=.55) and was characterized by significant and substantial loadings (i.e., ranging from .57 to .81). The second order CFA model for the MBC is presented in Figure 1.

Figure 1

Correlation between the two higher factors, Externalizing and Internalizing problems was r=.32. The estimated correlations between the Externalizing factor and Rules breaking, Hyperactivity/impulsivity, Lack of attention factors were: r=.95, r=.98 and r=.85 respectively. In addition, the correlations between the Internalizing domain and the four factors: Low energy, Stereotyped behavior, Lack of social interaction and Lack of self regulation, were: r=.81, r=.80, r=.95 and r=.97 respectively. Correlations between the seven factors and the two broadband, Externalizing and Internalizing problems factors are presented in Table 2.

Table 2

Model Fit

The results from the CFA on the second half of the data (*N*=415) and examination of the fit indexes supported the adequacy of the reduced model fit. More specifically, the CFI was .96, the NFI was .95 and the GFI was .92. In addition, the Root Mean of Approximation (RMSEA) was .072 and the Standardized Root mean Square Residual (SRMS) value was .054, which according to Hu and Bentler (1999) are good scores as values less than .80 for RMSEA and SRMS indices are acceptable and indicate good fit for the factor model.

Internal consistency

Alpha values for all the subscales were excellent suggesting that the list was homogenous in content. More specifically, for the factor Rules breaking (7 items), alpha value was .95, for factor Low energy (4 items), alpha=.82, for factor Stereotyped behavior (2 items), alpha=.85, for factor Hyperactivity/Impulsivity (14 items), alpha=.95, for factor Lack of attention (10 items), alpha=.95, for factor Lack of social interaction (10 items), alpha=.94 and finally for factor Lack of self regulation (12 items), the alpha coefficient was .91. In addition, for the externalizing scale (31 items), alpha coefficient was .93, and for the internalizing scale (28 items) the coefficient alpha was .91.

Reproducibility

Intraclass correlation coefficients (ICC) for each of the seven subscales were calculated separately. For the Rules breaking scale the ICC was: .87, for the Low energy factor was .78, for the Stereotyped behavior was .82, for factor Hyperactivity/Impulsivity was .90, for the factor Lack of attention was .89, for factor Lack of social interaction was .85 and finally for the factor Lack of self regulation the ICC coefficient was .83. For the externalizing scale the ICC coefficient was .87, and for the internalizing scale the ICC coefficient was .81. All correlation were highly significant at level p<.001 (see Table 3).

Inter-rater reliability

The ICC coefficients were significant at p<.001, ranging from .75 (Low energy) to .91 (Lack of attention). More specifically, for the scales of the Externalizing domain the inter-rater agreement was for the Rules breaking factor ICC=.88, for the Lack of attention factor ICC=.91 and for the Hyperactivity/Impulsivity factor ICC=.88. In addition, lower but statistical significant correlations were assessed for Internalizing factors. The correlation agreement for the Low energy factor was ICC=.75, for the Stereotyped behaviors was ICC=.85, for the Lack of Social interaction was ICC=.74, and for the Lack of self regulation factor was ICC=.81.In addition, ICC coefficient for the externalizing scale, was .78, and for the internalizing scale the ICC coefficient was .71. Internal consistency, test-retest and inter-rater reliability coefficients presented in Table 3.

Table 3

Summary of study two

The aim of study two was to investigate the key psychometric properties, namely: the structure validity, the internal consistency, the temporal stability and the inter-rater agreement, of

a new scale (i.e., the MBC) for the assessment of emotional and behavioral problems in elementary children by the physical educator teacher. A series of CFAs revealed a second order model with two (Externalizing and Internalizing) broadband factors and seven problems scales. The internal consistency was high for each scale suggesting that the list was homogenous in content. In addition, both the reproducibility and the inter-rater agreement were excellent suggesting that the MBC for children is an instrument with high temporal stability and high correlation agreement when used by physical educators in school settings.

General Discussion

A new instrument for physical education teachers

Students with EBD include a wide range of children and youth, including those with externalizing and internalizing behavior problems (Morris et al., 2002; Walker, Ramsey, & Gresham, 2004). These students often have broad-based needs because of their behavioral, social, and academic deficits which often do not improve over time (Lane, 2007; Mattison, Hooper, & Glassberg, 2002; Nelson et al., 2004). Because not all students with EBD will necessarily require special education it is very important that educators and school administrators, be prepared to implement systematic screening efforts to identify students who may show early signs of EBD. As such, the first step is to implement systematic screening tools to identify students who might benefit from more focused supports (Lane, 2007).

Physical educators may take a prominent position in the screening of children as they have the advantage of observing the child within a peer group, allowing these experts in movement situations to distinguish between maladaptive and typical age-related behaviors. The present study fills an important gap in the literature as physical educators lack a practical and reliable instrument for detecting children with behavioral and emotional problems on the basis of their motor related behaviors, despite the fact that a lot of useful information could be obtained through observation during physical education lessons in school settings or free play situations (Mol Lous et al., 2002).

The aim of study one was the development of a preliminary version of the MBC for children using reports from Greek physical educators about problematic student's behaviors and official psychiatric diagnostic criteria. The present study shows that PE teachers can observe deviant or problematic behavior among their students. Analyzing their reports it was obvious that they were able to generate a wide range of deviant behaviors, many of which resembled in

content with items coming from diagnostic frameworks. However, PE teachers tend to focus more on externalizing than on internalizing problematic behavior as the majority of their reports were motor related behaviors describing aggressiveness and disobedience exhibited by young children during lessons procedures. One possible explanation is that these types of behaviors are more difficult to deal with as they clearly disturb the class management, place demands on physical educators' management skills and are the most significant barrier to effective teaching (Siedentop & Tannehill, 2002; Vogler & French, 1983; White & Bailey, 1990).

The results of the sorting task indicated that the Physical Education teachers were able to discern patterns of similar behavior among the items, and a meaningful hierarchical cluster solution was revealed. Meaningful subclusters were also obtained from the cluster analysis and the hierarchical structure resembled the well-known distinction between externalizing and internalizing problem behavior. Based on the cluster solution and giving priority to the items from physical educators' reports, a number of items from each subcluster were selected by three experts for the preliminary version of MBC for children.

Psychometric properties of the MBC for children

Study two aimed to evaluate the key psychometric properties of the new instrument namely, structure validity, internal consistency, reproducibility, and inter-rater agreement. Results from a series of CFAs established a second order model with two (Externalizing and Internalizing) broadband domains and seven problems scales (Rules breaking, Hyperactivity/Impulsivity, Lack of attention, Low energy, Stereotyped behaviors, Lack of Social interaction, and Lack of self regulation). Items selection and items reduction per scale were based on statistical techniques (i.e., factor loadings, correlated uniqueness) and the selected items reflected areas that are important to the target population that is being studied. Therefore, the target populations (elementary students) as well as the target experts (physical education teachers) were involved during item selection.

Results from the internal consistency revealed high and significant values for each problem scale suggesting that the list was homogenous in content. Results from test-retest study support evidence that the list is characterized by satisfactory short-term stability. The level of correspondence between the test and retest was significant, but yet, it should be noted that the time lapse between the two administrations was relatively short (two weeks) and future research should verify whether the MBC results remain stable over longer period of time. In addition, the

inter-rater agreement was significant for all the problems scales, but the higher correlations agreements between the two physical educators were on the externalizing problems scales and especially on Rules breaking and Hyperactivity/Impulsivity scales, when the lowest agreement were noticed on the Lack of social interaction scale and the Low energy scale. The different educational settings between the two observers could partly justify the lower agreement on these Internalizing problem scales as social interaction and decreased activity are difficult to be observed at classroom settings.

An issue that disserves to be discussed is the incremental validity of MBC as a new instrument in the Greek culture. The fact that MBC for children was developed based on a theoretical and procedural framework derived from pilot studies in Greece makes it more ecological valid and appropriate in identifying culture specific aspects of a construct (Tsaousis & Georgiadis, 2009). Furthermore, the instrument contains items that were derived from the reports of Greek physical educators and as is demonstrated in the current study, the MBC for children has sound psychometric properties.

Limitations and future research

Participants were typical Greek elementary school-aged children. Findings may differ with a more diverse sample of clinical populations from psychiatric centers or from Special Elementary schools. For this reason, and in order to establish the discriminant validity of the list, future research studies using clinical samples of children are needed to investigate the ability of the instrument to discriminate clinical and controls with accuracy. Future research efforts are also needed to examine the concurrent validity and the ability of the Motor Behavior Checklist for children to provide similar ratings with valid and reliable measures in children's psychopathology.

Conclusions

Taking into consideration that early identification for emotional and/or behavioral problems can help to minimize the long-term harm of mental disorders and reduce the overall healthcare burden and costs (Aos et al, 2004), the MBC for children could be used for various educational purposes including research projects and intervention programs. The findings of this study are quite encouraging for the future use of MBC for children in the Greek population. Psychometric results supported the model suggesting that MBC for children is a new instrument homogenous in content, with high temporal stability and high correlation agreement that can provide useful and reliable

ratings on behavioral and emotional problems in children when used by physical educators in school settings. More specifically, the MBC can be used in estimating children's problematic motor related behaviors and the effectiveness of intervention programs in education aiming to reduce inappropriate behaviors. The information provided by the MBC may contribute to physical educators in developing class management techniques and assess the effectiveness of their interventions with a pre-post administration of the instrument. Moreover, one very important issue connected with special education settings, is that the information provided, when assessing children's deviant behaviors in a valid and systematic way within school settings, may help physical education teachers to decide about the referral or not of children for further diagnostic evaluation.

References

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

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232.

Aos, S., Lieb, R., Mayfield, J., Miller, M., & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth.* Olympia: Washington State Institute for Public Policy.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Bullis, M., & Yovanoff, P. (2006). Idle hands: Community employment experiences of formerly incarcerated youth. *Journal of Emotional and Behavioral Disorders*, *14*, 71–85.

Boomsma, D. I., van Beijsterveldt, C. E. M., & Hudziak, J. J. (2005). Genetic and environmental influences on anxious/depression during childhood: A study from the Netherlands Twin Register. *Genes, Brain and Behavior, 4,* 466-481.

Bruininks, R., & Bruininks, B. (2005). *Bruininks-Oseretsky Test of Motor Proficiency*-2nd edition manual. Minneapolis, MN: NCS Pearson.

Conners, C. K. (2000). *Conners' Rating Scales-Revised Technical Manual.* North Tonawanda, New York: Multi Health Systems.

DiStefano, C., & Kamphaus, R.W. (2007). Development and Validation of a Behavioral Screener for Preschool-Age Children. *Journal of Emotional and Behavioral Disorders, 15*, 93-102.

Drabick, D. A. G., Gadow, K. D., & Loney, J. (2008). Co-occurring ODD and GAD symptom groups: Source-specific syndromes and cross-informant comorbidity. *Journal of Clinical Child and Adolescent Psychology 37*, 314-326.

DuPaul, G. J. (1998). *ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation*. New York, NY: Guilford Press.

Epstein, M. H. (2004). *Behavioral and Emotional Rating Scale– Second Edition.* Austin, TX: PRO-ED.

Flanagan, K. S., Bierman, K. L., & Kam, C. M. (2003). Identifying at-risk children at school entry: The usefulness of multibehavioral problem profiles. *Journal of Clinical Child and Adolescent Psychology, 32,* 314-326.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 6,* 141-148.

Gadow, K. D., Drabick, D. A. G., Loney, J., Sprafkin, J., Salisbury, H., Azizian, A., et al. (2004). Comparison of ADHD symptom subtypes as source-specific syndromes. *Journal of Child Psychology and Psychiatry, 45*, 1135-1149.

Greenbaum, P. E., Dedrick, R. F., Prange, M. E. & Friedman, R. M. (1994). Parent teacher and child ratings of problem behaviors of youngsters with serious emotional disturbances. *Psychological Assessment, 6* (2), 141-148.

Gresham, F. M. & Elliott, S. N. (1990). *Social Skills Rating System manual.* Circle Pines, MN: American Guidance Service.

Henderson, S. E., & Sugden, D. A. (2007). *Movement Assessment Battery for Children.* Second Edition. London, The Psychological Corporation.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Hudziak, J. J., Derks, E. M., Althoff, R. R., Copeland, W. & Boomsma, D. I. (2005). The genetic and environmental contributions to oppositional defiant behavior: A multi-informant twin study. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 907-914.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8.11.* Chicago: Scientific Software International.

Kashani, J. H., Allan, W. D., Beck, N.C., Bledsoe, Y., & Reid, J. C. (1997). Dysthymic disorder in clinically referred preschool children. *Journal of the American Academy of Child and Adolescents Psychiatry, 36*, 1426-1433.

Kauffman, J. M., & Landrum, T. (2009). *Characteristics of emotional and behavioral disorders of children and youth* (8th ed.). Columbus, OH: Merrill.

Klassen A., Miller A., & Fine S. (2004). Health-Related Quality of life in children and adolescents who have a diagnosis of Attention-Deficit/Hyperactivity Disorder. *Pediatrics, 114*, 541-547.

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent, and teacher reports. *Journal of Child Psychology and Allied Disciples, 34(6),* 991-1006.

Kulinna, P.H., Cothran, D.J. &. Regualos, R. (2003). Development of an instrument to measure student disruptive behavior. *Measurement in Physical Education and Exercise Science 7,* 25–41.

La Greca, A.M. (1999). *Social Anxiety Scale for children-Revised (SASC-R),* Manual. Department of Psychology, University of Miami, Coral Gables.

Landrum, T. J., Tankersley, M., & Kauffman, J. M. (2003). What is special about special education for students with emotional and behavioral disorders? *Journal of Special Education, 37,*148–156.

Lane, K. L. (2003). Identifying young students at risk for antisocial behavior: The utility of "teachers as tests." *Behavioral Disorders*, *28*, 360–389.

Lane, K. L. (2007). Identifying and supporting students at risk for emotional and behavioral disorders within multi-level models: Data driven approaches to conducting secondary interventions with an academic emphasis. *Education and Treatment of Children*, *30*, 135–164.

Marsh, H. W., Ellis, L., Parada, L., Richards, G. & Heubeck, B. G. (2005). A short version of the Self Description Questionnaire II: Operationalzing criteria for short-form *evaluation with new applications of confirmatory factor analyses. Psychological Assessment, 17,* 81-102.

Martin, N., Scourfield, J., & McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *British Journal of Psychiatry, 180,* 260-265.

Mattison, R. E., Hooper, S. R., & Glassberg, L. A. (2002). Three-year course of learning disorders in special education students classified as behavior disordered. *Journal of the American Academy of Child and Adolescent Psychology*, *41*, 1454–1461.

Messick, S. (1995). Validity of psychological measurement: Validation of inferences from a person's responses and performance in scientific inquiry into score meaning. *American Psychologist, 50*, 741 – 749.

Mol Lous, A., Wit, C. A. M., De Bruyn, E. J., & Riksen-Walrawen J. M. (2002). Depression markers in young children's play: a comparison between depressed and non

depressed 3 to 6 years old in various play situations. *Journal of Child Psychology and Psychiatry, 43* (8), 1029-1038.

Morris, R. J., Shah, K., & Morris, Y. P. (2002). Internalizing behavior disorders. In K. *Lane, F. Gresham, & T. O'Shaughnessy (Eds.), Interventions for children with or at risk for emotional and behavioral disorders* (pp. 223–241). Boston: Allyn & Bacon.

Nelson, N. J., Babyak, A., Gonzalez, J., & Benner, G. J. (2003). An investigation of the types of problem behaviors exhibited by K–12 students with emotional or behavioral disorders in public school settings. *Behavioral Disorders*, *28*, 348–359.

Quay, H. C., & Peterson, D. R. (1987). *Manual for the Revised Behavior Problem Checklist.* Coral Gables, FL: University of Miami, Department of Psychology.

Reynolds, C. R., & Kamphaus, R. W. (2004). *BehaviorAssessment System for Children– Second Edition (BASC-2).* Circle Pines, MN: American Guidance Service.

Rosenberg, S., & Jones, R. (1972). A method for investigating and representing a person's implicity theory of personality: Theodore Dreise's view of people. *Journal of Personality and Social Psychology, 22*, 372-386. doi:10.1037/h0032891.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, *10*, 489–502.

Siedentop, D., & Tannehill, D. (2002*). Developing teaching skills in physical education (4th ed.).* Mountain View, CA: Mayfield Publishing.

Spielberger, C. D. (1973). *Manual for the State-Trait Anxiety Inventory for Children.* Palo Alto, CA: Consulting Psychologists Press.

SPSS 15.0. (2006). *Statistical Package for the Social Sciences. Command Syntax Reference.* SPSS Inc., Chicago III.

Tsaousis, I. & Georgiadis, S. (2009). The Greek Personality Adjective Checklist. *European Journal of Psychological Assessment, 25* (3), 164-174.

Ulrich, D. (2000). *Test of gross motor development. Examiner's manual (2nd ed.).* Austin, Texas: PRO-ED.

Verhulst, F.C., & Akkerhuis, G.W. (1989). Agreement between parents' and teachers' ratings of behavioral/emotional problems of children aged 4-12. *Journal of Child Psychology and Psychiatry, 30*(1), 123-136.

Vogler, E.W., & French, R.W. (1983). The effects of a group contingency strategy on behaviorally disordered students in physical education. *Research Quarterly for Exercise and Sport, 54*, 273-277.

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T. & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 45*, 135-170.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices*. Belmont, CA: Wadsworth.

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi,W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbance receiving special education services. *Journal of Emotional and Behavioral Disorders*, *13*, 79–96.

Wagner, M., & Davis, M. (2006). How are we preparing students with emotional disturbances for the transition to young adulthood? Findings from the National Longitudinal Study-2. *Journal of Emotional and Behavioral Disorders*, *14*, 86–98.

Wagner, M., Newman, L., Cameto, R., Levine, P., & Garza, N. (2006). *An overview of findings from Wave 2 of the National Longitudinal Transition Study-2 (NLTS2)*. Menlo Park, CA: SRI International.

White, A.G., & Bailey, J.S. (1990). Reducing disruptive behaviors of elementary physical education students with sit and watch. *Journal of Applied Behavior Analysis, 23*, 353-359.

World Health Organization. (1992). *The IDC-10 classification of mental and behavioral disorders: Diagnostic criteria for research*. Geneva: Author.

Zigmond, N. (2006). Twenty-four months after high school: Paths taken by youths diagnosed with severe emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *14*, 99–107.