Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Research in Developmental Disabilities 33 (2012) 1343–1351

Contents lists available at SciVerse ScienceDirect

Research in Developmental Disabilities

Agreement among physical educators, teachers and parents on children's behaviors: A multitrait–multimethod design approach

Maria Efstratopoulou [a], Rianne Janssen [b], Johan Simons [a,*]

[a] Faculty of Kinesiology and Rehabilitation Sciences, Research Centre for Adapted Physical Activity and Psychomotor Rehabilitation, Catholic University of Leuven, Belgium

[b] Faculty of Psychology and Educational Sciences, Centre for Educational Effectiveness and Evaluation and Research Group of Quantitative Psychology and Individual Differences, Catholic University of Leuven, Belgium

ARTICLE INFO

Article history:
Received 6 March 2012
Accepted 7 March 2012
Available online

Keywords:
Convergent validity
Multitrait–multimethod
Emotional and behavioral problems
Children
Physical education

ABSTRACT

The study examines the agreement among raters on children's problematic behaviors. A multitrait–multimethod (MTMM) matrix was applied to a normative sample of elementary school-aged children ($N = 841$). The participants were rated by their physical educators, using the Motor Behavior Checklist for children (MBC; Efstratopoulou, Janssen, & Simons, 2012). Teachers and parents rated the same students using the Teacher Report Form (TRF; Achenbach, 1991b), the Child Behavior Checklist (CBCL; Achenbach, 1991a) and the ADHD Rating Scale-IV (DuPaul, Power, Anastopoulos, & Reid, 1998). The resulting matrix revealed significant correlations for the Rules Breaking, Lack of Attention, Hyperactivity/Impulsivity, Lack of Social interaction problem scale and for the Internalizing, Externalizing and Total scores. Convergent validity of the specific MBC subscales was supported by significant correlations with the corresponding subscales of TRF, CBCL and ADHD Rating Scale-IV. Findings underscore the importance of taking child's settings and observer influences into account and suggest that MBC is a new promising instrument that can provide valid ratings on externalizing behavior and social problems in children when used by physical educators in school settings.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Detection efforts for students with emotional, behavioral and developmental disorders are particularly critical during the early educational years, when these children are most amenable to change in behavioral, social, and academic arenas and before experience negative outcomes within and beyond the school setting (Landrum, Tankersley, & Kauffman, 2003; Lane, 2003; Volkmar, Lord, Bailey, Schultz, & Klin, 2004).

Given the costs associated with these disorders to students themselves, their families, and society as a whole, it is not surprising that reducing the incidence of emotional, behavioral and developmental disorders through systematic screening and comprehensive intervention efforts is a growing area of interest in educational research (Kauffman & Landrum, 2009; Lane, 2007; Nelson, Babyak, Gonzalez, & Benner, 2003).

* Corresponding author at: Faculty of Kinesiology and Rehabilitation Sciences, Catholic University of Leuven, Belgium Tervuursevest 101, 3001 Heverlee, Belgium.
E-mail address: johan.Simons@faber.kuleuven.be (J. Simons).

0891-4222/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ridd.2012.03.015

Author's personal copy

1.1. Agreement among rating sources

Despite the usefulness of rating instruments for screening children's deviant behaviors, the relatively modest agreements among rating sources it is a problem concerning the validity of the information and the importance of context or setting effects on children's behavior. Information on children's behavior can be gathered by a number of informants who each have their own point of view. Parents can observe their child in a wide range of situations; nonetheless, information from the parents is not always reliable. The accuracy of parents as raters may vary depending on such factors as education, the amount of stress associated with the child's behaviors, and hidden agenda's that parents may have when rating a child (De Los Reyes & Kazdin, 2005). Some parents tends to follow a pattern of idealized expectations and cultural stereotypes, some may be very sensitive or may have a low threshold for certain behaviors and will exaggerate symptoms, whereas other parents may underreport deviant or troublesome child behaviors.

Modest agreement is the norm for different informants' ratings of a given child's functioning (Achenbach, McConaughy, & Howell, 1987; Achenbach, 1991a), raising concerns about the relative validity of any single source of information and creating measurement obstacles for both research and clinical endeavors. Mental health professionals have distinct opinions about the relative value of different informants for behavioral criteria. The frequency, base rate, and conspicuousness of behaviors 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., Drabick, Gadow, & Loney, 2008; Gadow et al., 2004). In general, concordance has been found to be higher when informants have similar relationships with the children being rated than when raters represent different roles (Achenbach et al., 1987; Greenbaum, Dedrick, Prange, & Friedman, 1994). There is also reason to believe that children's presentation of behavior problems varies across different settings such as the home or school (Kazdin & Kagan, 1994; Kolko & Kazdin, 1993). Additionally, cross-informant agreement may vary considerably depending on item content: Teachers may be more sensitive to disruptive behavior and parents more to depression or anxiety (Abikoff, Courtney, Pelham, & Koplewicz, 1993). On the other hand, teachers are more likely to agree with other teachers about depressive symptoms or overall levels of depressive and aggressive behaviors (Epkins, 1995; McDermott, 1994), suggesting that there may be more cross-situational continuity in these behaviors. Teacher reports also may be more specific for these symptoms versus disruptive or attentional problems, in as much as teachers appear less likely to over report internalizing versus disruptive problem behaviors (Abikoff et al., 1993). Another view is that agreement should be greater about externalizing behaviors, whereas internalizing behaviors might be more difficult to observe and less disruptive to family or classroom functioning and therefore less likely to attract the attention of adult informants (Achenbach et al., 1987; Kolko & Kazdin, 1993).

Clinicians and researchers generally perceive children and youth self-report as the least useful source of behavior ratings pertaining to hyperactivity, inattention, and oppositional behaviors; whereas both youths and caregivers are preferred to teachers as sources of information about internalizing problems (Youngstrom, Loeber, & Stouthamer-Loeber, 2000).

1.2. The role of physical education teachers in the assessment procedure

Several studies suggest 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 physical education classes, team games and during standardised play procedures (Kashani, Allan, Beck, Bledsoe, & Reid, 1997; Mol Lous, Wit, De Bruyn, & Riksen-Walrawen, 2002). However, there are only a few instruments that use 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, problems in concentration, balance and levels of motor skill development as part of psychological test batteries, for making decisions about educational placement, developing and evaluating intervention programs.

The DSM-IV criteria for Disruptive behavior disorders (DBDs) include several items related to motor characteristics (American Psychiatric Association, 2000). During physical activities, children with ADHD exhibit age-inappropriate features of hyperactivity, excessive impulsivity, problems in lateralization, and are often left-handed (Corrigan, 2003; Reid & Norvilitis, 2000). Children with conduct problems at educational settings deviate from school and social principles, rules and regulations; display delinquent behavior, difficulties in social relationships, aggressiveness, combustible disobedience, anger, lack in empathy or concern for others, misperception of the intent of others in ambiguous social situations, lack in guilt or remorse, and low self-esteem (Dodge, 1993).

The importance of examining the interaction of peer relationships and other social relationships in the physical activity context, and the value of using the physical activity setting to promote quality peer relationships have been established (Smith, 2003). Motor related behaviors observed in children with developmental disorders include physical aggression, self-injury, property destruction, stereotyped behaviors, and tantrums which are highly disruptive to classroom, community, and home environments and without intervention, they are more likely to increase than improve (Horner, Carr, Strain, Todd, & Reed, 2002). During physical activities, children with ASD, indicate stereotyped and repetitive motor mannerisms, impairments of facial expression, postures, and gestures, and are often characterized as clumsy and as having problems in motor coordination (Berkeley, Zittel, Pitney & Nichols, 2001; Piek & Dyck, 2004).

Author's personal copy

Physical education teachers 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 in several different settings (inside or outside the classroom, at the playground or at the school-yard). However, there is a lack of literature concerning the investigation of the added value of the information provided by physical educators on children's emotional, behavioral and social problems in school settings.

1.3. Hypotheses of the current study

The study aimed to investigate the convergent validity of the Motor Behavior Checklist (MBC; Efstratopoulou et al., 2012) using data from a typical elementary school-aged sample. A multitrait–multimethod (MTMM) research design (Campbell & Fiske, 1959) was used to examine the validity of specific problem scales of the MBC. Correlations among scores on related problem scales were examined. Using data from three different sources of information, (i.e., physical educators, teachers and parents), and ratings on four behavioral assessment instruments (MBC, Efstratopoulou et al., 2012; TRF, Achenbach, 1991b; CBCL, Achenbach, 1991a; ADHD Rating Scale-IV, DuPaul et al., 1998) the agreement among different raters on specific problem scales was also investigated. The following hypotheses were tested:

(i) There are significant correlations among corresponding scales of the instruments used. More specifically, convergent relationships are expected for the Rules Breaking problem scale of MBC and the Delinquent Behavior scales of TRF and CBCL. In addition, convergent relationships are also expected on Attention problem scales, Social Problem scales, Externalizing, Internalizing, and Total scores.

(ii) The multitrait correlations, discussed above, are significantly higher when measured by raters in school settings (e.g. teachers; physical educators) than when measured across settings (e.g. teachers; parents).

2. Method

2.1. Participants

The data derived from 35 typical Greek elementary schools widely spread across the country 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 data analysed were collected from a randomly selected sample ($N = 841$) of elementary school-aged children. The overall sample consisted of 421 (50, 1%) girls and 420 (49, 9%) boys, ranging from 6 to 11 years ($M = 8.4$ years, $SD = 1.7$ years) and they had the Greek nationality (99%).

2.2. Assessment instruments

2.2.1. Motor Behavior Checklist (MBC) for children

The Motor Behavior Checklist for children (MBC; Efstratopoulou et al., 2012) is a scale designed to be completed by the physical education teacher who knows the child well enough to rate his/her motor behavior. Raters are asked to observe the child during physical education classes and free play situations and to rate each behavior on a Likert scale ranging from ''never'' (0) to ''almost always'' (4). The MBC for children consists of 59 motor related behavior items included in two broadband factors (Externalizing and Internalizing) and seven problems scales: Rules Breaking (7 items), Hyperactivity/Impulsivity (14 items), Lack of Attention (10 items), Low energy (4 items), Stereotyped Behaviors (2 items), Lack of Social interaction (10 items), and Lack of Self regulation (12 items). The internal consistency (ranging from $\alpha = .82$ to $\alpha = .95$), the reproducibility (ranging from $ICC = .85$ to $ICC = .90$) and the interrater agreement (ranging from $ICC = .75$ to $ICC = .91$) are excellent suggesting that the MBC for children is an instrument homogenous in content, with high temporal stability and high correlation agreement. In addition, the discriminant validity of the list was established in a study using clinical samples of elementary school-aged children (Efstratopoulou et al., 2012).

2.2.2. Child Behavior Checklist (CBCL) and Teacher Report Form (TRF)

One of the most popular approaches to measure childhood behavior problems has been to use rating scales that are completed by either parents or teachers. The Child Behavior Checklist (CBCL; Achenbach, 1991a) and Teacher Report Form (TRF; Achenbach, 1991b) are among the most widely used measures of children's emotional and behavioral problems in both clinical and research settings. The items measure three broad-band scales: Internalizing, Externalizing, and Total Problems, and eight syndrome scales: Withdrawal, Somatic Problems, Anxiety/Depression, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior (Achenbach, 1991a, 1991b). The items on both CBCL and TRF, were rated as Not True (0), Somewhat or Sometimes True (1), or Very True or Often True (2), and summed to yield (a) eight syndrome scale scores, (b) six DSM-oriented scale and (c) broad-band scale scores (including internalizing and externalizing total scores). With well-established normative data and standardized clinical cutoffs, the instruments have demonstrated strong psychometric properties (Achenbach, 1991a, 1991b; Chen, Faraone, Biederman, & Tsuang, 1994; Drotar, Stein, & Perrin, 1995).

Author's personal copy

Table 1
Participant's characteristics.

Variables		Children ($N = 841$)	Teachers ($N = 210$)	Physical educators ($N = 62$)	Parents ($N = 841$)
Gender	Male	49.9%	37.1%	43.5%	20.7%
	Female	50.1%	62.9%	56.5%	79.3%
Nationality	Greek	99.4%	100%	100%	99.4%
Age in years[a]		8.4 (1.7)	38.7 (8.7)	39.4 (6.2)	36.5 (9.3)
Experience in years[a]			11.4 (4.2)	7.2 (3.4)	

[a] Standard deviation in parentheses.

2.2.3. ADHD Rating Scale-IV

The ADHD Rating Scale-IV (DuPaul et al., 1998) is a reliable and easy-to-administer instrument both for diagnosing ADHD in children and adolescents and for assessing treatment response. Containing 18 items, the scale is linked directly to DSM-IV (American Psychiatric Association, 2000) diagnostic criteria for ADHD. The manual provides two versions of the scale: a parent questionnaire on home behaviors, and a teacher questionnaire on classroom behaviors. The items on ADHD scale, were rated as; almost never (0), rarely (1), many times (2), very often (3) and summed to yield (a) a total score and (b) separate scores on attention and on hyperactivity/impulsivity items.

2.3. Procedure

All participants and their legal guardians underwent standardized Institutional Review Board-approved notice of privacy and consent procedures. The study was in line with the guidelines given by the Research Ethics Board of the K.U. Leuven and was approved by the Pedagogy Department of Greek Ministry of Education. The participant's classroom teachers ($N = 210$) were asked to select in a random way, four children (two boys and two girls) from each grade and to rate them on the TRF (Achenbach, 1991b) and ADHD Rating Scale-IV (DuPaul et al., 1998). The physical education teachers of the schools ($N = 62$) were asked to rate the same students using the Motor Behavior Checklist (MBC) for children. In addition, the parents of the participant's children were asked to fill in the parent's version of Achenbach's test (CBCL, Achenbach, 1991a) and the parent's form of the ADHD Rating Scale-IV (DuPaul et al., 1998). The parents were informed about the aim of the study and the assessment procedure, by the research assistant and the classroom teacher of the school. It was clarified that the children were selected in a random way, and that the data will be treated as anonymous and confidential. An overview of participant's characteristics is in Table 1.

2.4. Statistical analysis

2.4.1. The multitrait–multimethod analysis

The multitrait–multimethod MTMM analysis (Campbell & Fiske, 1959) was invented to investigate the convergent validity of scales on the basis of their inter correlations and has proven to be one of the most powerful tools for detecting trait, method, and error components of a measurement. Multimethod research designs are nowadays preferred to single method designs in almost all areas of psychological research (Eid & Diener, 2006) as is proven by an increasing number of studies include investigation of social adaptation, and deviant behavior (Allen, Porter, McFarland, Marsh, & McElhaney, 2005); childhood depression and anxiety (Tram & Cole, 2006); parenting effects on the mental health of bereaved children (Kwok et al., 2005); temperament in early childhood and aggressive behavior in children (Majdandzic & van den Boom, 2007; Ostrov & Crick, 2007).

More sophisticated MTMM models (e.g., Eid, 2000; Eid, Lischetzke, & Nussbeck, 2006; Marsh, 1989; Marsh & Grayson, 1995) may serve to answer research questions related to measurements and the convergence relations between ratings measured by different raters or different methods. Tables of correlations arranged to facilitate the interpretation of the assessment of validity are created. The method assumes that each of several concepts (called *traits* (Campbell & Fiske, 1959) is measured by each of several instruments (e.g., MBC, Efstratopoulou et al., 2012; TRF, Achenbach, 1991b; CBCL, Achenbach, 1991a and ADHD Rating Scale-IV, DuPaul et al., 1998) using data from different raters. Campbell and Fiske (1959) suggested that validity coefficients should be higher than values lying on its column and row in the same heteromethod block and these values should be significantly different from zero and sufficiently large to encourage further examination of validity. Six parallel or like-named categories from TRF and CBCL were selected for this study. Because the measures were of the same trait or concept, it was expected to be strongly correlated and these values considered to be monotrait-heteromethod correlations. Since the MTMM is organized into method blocks, there is one validity diagonal in each method block. The Statistical Package for the Social Sciences (SPSS 15.0, 2006) was used for the analysis of the data.

3. Results

The convergent validity of the Motor Behavior Checklist problem scales (MBC, Efstratopoulou et al., 2012) was examined through the correlation coefficients along the validity diagonals. A multitrait–multimethod-correlation matrix (Campbell &

Author's personal copy

Table 2
Correlation Matrix: Motor Behavior Checklist (MBC), Teacher Report Form (TRF), and Child Behavior Checklist (CBCL) subscales.

	Physical Educators *MBC*						Teachers *TRF*						Parents *CBCL*					
MBC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.Rules	-																	
2.Attention	.42**	-																
3.Social	.17*	.36**	-															
4.Externalizing	.78**	.85**	.21*	-														
5.Internaling	.13*	.15*	.79**	.19*														
6.Total Scale	.76**	.83**	.62**	.87**	.63**	-												
TRF																		
7.Delinquent	**.58****	.37**	.03	.45**	.09	.35**	-											
8.Attention	.36**	**.56****	.10*	.45**	.12*	.41**	.55**	-										
9.Social	.24*	.19*	**.33***	.16*	.27*	.21*	.48**	.67**	-									
10.Externalizing	.43**	.33**	.10*	**.58****	.17*	.41**	.91**	.81**	.71**	-								
11.Internalzing	.06	.07	.27*	.10*	**.46****	.27*	.33**	.36**	.61**	.29**	-							
12.Total Scale	.37**	.28*	.27**	.41**	.22*	**.52****	.84**	.91**	.78**	.89**	.33**	-						
CBCL																		
.113.Delinquent	**.52****	.23**	.04	.41**	.06	.32**	**..47****	.45**	.38**	.44**	.24*	.46**	-					
.114.Attention	.27*	**.52****	.13*	.42**	.11*	.37**	.50**	**.46****	.40**	.54**	.16*	.56**	.84**	-				
15.Social	.10**	.09	**.28***	.18*	.24*	.19*	.38**	.35**	**.33****	.39**	.22*	.40**	.64**	.56**	-			
16.Externalizing	.33**	.31*	.11*	**.52****	.13*	.39**	.51**	.48**	.42**	**.48****	.16*	.51**	.87**	.74**	.70**	-		
17.Internalizing	.14**	.17*	.27*	.10*	**.34***	.17*	.33**	.21*	.24**	.27**	**.31****	.32*	.67**	.47**	.65**	.57**	-	
18.Total Scale	.30*	.26*	.29*	.39**	.23*	**.48****	.52**	.45**	.42**	.52**	.27*	**.42****	.90**	.75**	.83**	.82**	.82**	-

*Correlation is significant at $p < .05$.
**Correlation is significant at $p < .01$.

Fiske, 1959) of all the inter-correlations between the six subscales and the three measurements (physical educators, teachers and parents) was generated, and the results are in Table 2. Convergent validities are the bold values.

Examination of the validity diagonals confirms the convergent validity of these MBC subscales. The subscales of MBC appear to confirm the initial hypothesis (i) that there are significant correlations among corresponding scales of the instruments used. More specifically, the problem scales of MBC for children (Rules Breaking, Lack of Attention, Lack of Social interaction, Externalizing, Internalizing and Total scale) demonstrated significant convergence with TRF (Achenbach, 1991b) scales (Delinquent, Attention problems, Social problems, Externalizing, Internalizing and Total scale) with coefficients of $r = .58$ ($p < .01$), $r = .56$ ($p < 01$), $r = .33$ ($p < 05$), $r = .58$ ($p < 01$), $r = .46$ ($p < 01$) and $r = .52$ ($p < 01$), respectively.

In addition, the mean scores of the six subscales of the MBC (Rules Breaking, Lack of Attention, Lack of Social interaction, Externalizing, Internalizing and Total scale) demonstrated positive relationships and significant convergent with the mean scores on the CBCL (Achenbach, 1991a) scales (Delinquent, Attention problems, Social problems, Externalizing, Internalizing and Total problem scale) with coefficients of $r = .52$ ($p < .01$), $r = .52$ ($p < .01$), $r = .28$ ($p < 05$), $r = .52$ ($p < .01$), $r = .34$ ($p < 01$), and $r = .48$ ($p < 01$), respectively. Concerning the examination of the initial hypothesis (ii) the results revealed that the correlations between the mean scores of physical educators and classroom teachers were significantly higher than between the mean scores of physical educators' and parents' ratings for all the subscales studied.

The convergent validity of the Lack of Attention and Hyperactivity/Impulsivity scales of MBC (Efstratopoulou et al., 2012) were also examined separately with ratings on ADHD Rating scales (DuPaul et al., 1998). A multimethod correlation matrix (Campbell & Fiske, 1959) of all the intercorrelations between the two subscales (Attention, Hyperactivity/Impulsivity) and the three measurements (physical educators, teachers and parents) was generated, and the results are in Table 3. Convergent validities are the bold values.

The results appear to confirm the hypothesis that there is a convergent relationship between the mean scores on Lack of Attention and Hyperactivity/Impulsivity problem scales when measured by different instruments (MBC and ADHD rating scales) and by different raters (e.g. physical educators, teachers, parents). More specifically, the Lack of Attention and Hyperactivity/Impulsivity problem scales of MBC demonstrated convergent relations with ADHD Rating scale-IV (teachers' version) with correlation coefficients of $r = .55$ ($p < .01$), and $r = .61$ ($p < .01$), respectively. The Lack of Attention and Hyperactivity/Impulsivity problem scales of MBC demonstrated also convergent relations with ADHD rating scale (parents' version) with correlation coefficients of $r = .43$ $p < 01$), and $r = .35$ ($p < .01$), respectively. With regard to the Attention and Hyperactivity/Impulsivity scales, the correlations found between physical educators and teachers were higher than between physical educators and parents for both scales.

Finally, concerning the inter correlations among the MBC subscales, the results indicated moderate and positive relationship of $r = .42$ ($p < .01$) between Rules Breaking and Lack of Attention scales, low and positive correlation of $r = .17$

Author's personal copy

Table 3
Correlation Matrix: Motor Behavior Checklist (MBC) and ADHD-Rating Scale-IV on attention and hyperactivity subscales.

Raters	Instrument scale	Physical Educators		Teachers		Parents	
		MBC Attention	*MBC* Hyperactivity impulsivity	*ADHD-RS-IV* Attention	*ADHD-RS-IV* Hyperactivity	*ADHD-RS-IV* Attention	*ADHD-RS-IV* Hyperactivity
Physical educators	*MBC* Attention	–					
	MBC Hyperactivity/Impulsivity	.78*	–				
Teachers	*ADHD-RS-IV* Attention	**.55****	.54**	–			
	ADHD-RS-IV Hyperactivity	.52**	**.61****	.78**	–		
Parents	*ADHD-RS-IV* Attention	.43**	.40**	**.62****	.51**	–	
	ADHD-RS-IV Hyperactivity	.31**	.35**	.50**	**.48****	.73**	–

** Correlation is significant at $p < .01$.

($p < .05$) between Lack of Social interaction and Rules Breaking and a moderate positive correlation of $r = .36$ ($p < 01$) between Lack of Attention and Lack of Social interaction scale. With regard to the Externalizing and Internalizing scales of MBC the correlation coefficient between the mean scores was $r = .19$ ($p < .05$).

4. Discussion

The purpose of this study was to investigate the convergent validity of the Motor Behavior Checklist (MBC; Efstratopoulou et al., 2012) using data from a typical elementary school-aged sample. The MTMM model (Campbell & Fiske, 1959) was used in order to examine the agreement among raters on children's problematic behavior. Based on the MTMM design we created tables of correlations to examine the validity diagonals between similar contracts (subscales) rated by different raters and we tested hypotheses concerning the relationships among problem scales.

The results indicated that the independent measurement methods significantly demonstrated convergence with all the scales studied. That is, the measurement methods did show a significant degree of convergent or concurrent validity with the behavioral problem scales of Rules Breaking, Lack of Attention, Hyperactivity/Impulsivity, Lack of Social interaction, Externalizing, Internalizing and Total Scale. As it was hypothesized, there was a convergent relationship between the Rules Breaking scale of MBC (Efstratopoulou et al., 2012) and the Delinquent scale of TRF (Achenbach, 1991b) and CBCL (Achenbach, 1991a), and between the Lack of Attention problem scale of MBC and the Attention scale of TRF and CBCL. The results indicated positive and significant correlations among the mean scores, with correlations being significantly higher between physical educators and teachers than between physical educators and parents. These results are consisted with literature indicating that concordance has been found to be higher when informants have similar relationships (e.g., teacher–physical educator) with the children being rated (Achenbach et al., 1987; Greenbaum et al., 1994) than when raters represent different roles (e.g., physical educators–parents).

Interestingly, for the Lack of Social interaction scale the correlations among the MBC and Social Problems scale of TRF and between the MBC the Social Problems scale of CBCL were significant but rather low although the similarity in content. One possible explanation for this low correlation could be the different settings in which the child was being rating. Even when both raters are educators there are differences concerning the motor behavior characteristics of a child that can be observed and recorded in school settings. For example, during physical education class and group play situations the social participation of children, as members of a team, is motivated and students are forced, by the nature of the lesson, to co-operate and to socialize. The child is more motivated to participate in group play situations and team games than during classroom lessons or at home and this could limited the possibility for a child to be rated as alone and isolated by peers. In addition, research in emotional and behavioral problems in children also indicates that Internalizing behaviors like social problems, anxiety and depression are more likely to be ignored (Abikoff et al., 1993) as often less apt to capture teacher's attention relative to students with Externalizing tendencies (Lane, 2007). These reasons could also justify the fact that the correlations between the MBC and TRF on behaviors like Delinquent and Attention problems and on Externalizing mean scores were significantly higher than the correlations between the MBC and the TRF on the Social problem scale and the Internalizing scores. In addition, the correlations were higher for the Externalizing scales (Rules Breaking and Lack of Attention) and Externalizing scores between the MBC and the CBCL than between the Lack of Social interaction scale and Internalizing scores of the MBC and the CBCL.

Author's personal copy

An issue that has to be discussed is the high correlation that was observed between the Delinquent and Attention problem scales on parents' ratings. Although that, based on existing research (Neuman et al., 2001; Newcorn et al., 2001; Moffit, 2003; Moffit, Caspi, Rutter & Silva, 2001), we expect significant associations on Attention and Rules Breaking scales, the high correlation ($r = .84$, $p < .01$), for parents' ratings indicates that maybe parents are not in a position to distinguish between inattention and delinquent children's behaviors. For the same scales (Rules Breaking and Lack of Attention) the correlations for physical educators (MBC) was positive and moderate ($r = .42$, $p < .01$) and for teachers (TRF, Delinquent and Attention problems) was also moderate ($r = .55$, $p < .01$). One possible reason is the fact that at home settings the Rules to follow are quite different than in school settings and this could possible means that parents perceived Delinquent or Rules Breaking behaviors in a quite different way than the teachers and the physical educators in school settings. In addition, behaviors like: the child is careless or cannot focus on tasks or cannot stay still, or even talk too much and interrupt others, may be perceived as disobedience from the parents' point of view.

With regard to the ADHD behaviors, as it was hypothesized, there were significant correlations on Lack of Attention and Hyperactivity/Impulsivity scales between MBC and ADHD Rating scales. High correlations were also observed between the Lack of Attention and Hyperactivity/Impulsivity scales on physical educators ratings on MBC and between Attention and Hyperactivity/Impulsivity scales on teachers' and parents' ratings which was up to a point expected by the fact that there is high coexistence of these ADHD behaviors. Finally, it must be noted that the MBC for children showed greater convergent validity with teachers' reports than with parents' report in all scales measured. Agreement between physical educators and classroom teachers was significantly higher than between physical educators and parents, suggesting that although parents are useful informants of their children's behaviors, educators have another point of view and are able to observe and rate the children it in a more systematic and comparative way.

A last step in the analysis of the results was the correlations revealed among the MBC subscales. It appears that although there were moderate correlations between Externalizing problem scales on MBC, the distinction between the Externalizing and Internalizing behaviors has some merit; although in clinical practice children with behavioral problems tend to exhibit a mixture of Internalizing and Externalizing problems (Roussos et al., 2001).

4.1. Implications, limitations and recommendations for future research

The findings of this study are quite encouraging for the future use of MBC (Efstratopoulou et al., 2012) for children in Elementary school-aged population. 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, Lieb, Mayfield, Miller, & Pennucci, 2004), the MBC for children could be used for various educational purposes including research projects and intervention programs and may contribute to physical educators in developing class management techniques and assess the effectiveness of their interventions with a pre–post administration. A further and more in depth accurate psychological assessment must follow this initial "screening" as the aim of MBC is not to provide a clinical diagnosis but to facilitate teaching procedure for physical education teachers in school settings and help them in their important decision to refer these students for further clinical evaluation.

Limitations of the present study include that the different informants described children behavior in partially overlapping contexts (e.g., parents report deriving from home functioning, teacher report from class functioning, and PE report from school-contexts). This lack of situational concordance matches how these instruments are generally used in research and clinical work, but it makes it difficult to discern whether disagreements reflect true differences in behavior across settings versus differences in perceptions among raters.

Another potential weakness of the research which could limit the generalizability of the results is the fact that the participants were typical Greek children in which possible differences on their motor related behaviors due to age or gender were not investigated in this study. Future studies could focus on possible differences on the way that physical educators rate deviant behavior (e.g., related to attention problems or rules breaking) for boys or girls or for first grade students in comparison with older students, during physical education classes in school settings.

Concerning the moderate agreement among rating sources on children's deviant behavior, future studies could focus on understanding why different raters often observe the same behaviors in different ways. Thus, future investigations should focus on why informants view children's behavioral problems similarly or differently and how this information could be valuable on the problems being assessed in a greater understanding of how to intervene in order to change these problematic behaviors within the school context.

In addition, as research studies suggest that teachers' characteristics, including self-efficacy, behavioral standards, stress and working burnout (Egye & Short, 2006), may influence their ratings, it could be interesting to take into consideration in a future study several types of teachers' and physical educators' characteristics in order to investigate from this point of view the agreement between these raters on students' behavior in school settings.

References

Abikoff, H., Courtney, M., Pelham, W. E., & Koplewicz, H. S. (1993). Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology, 21*, 519–533.

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. Burlington: University of Vermont Department of Psychiatry.

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

Author's personal copy

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.
American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th text rev.). Washington, DC: Author.
Allen, J. P., Porter, M. R., McFarland, C. F., Marsh, P. A., & McElhaney, K. B. (2005). The two faces of adolescents' success with peers: Adolescent popularity, social adaptation, and deviant behavior. *Child Development, 76*, 747–760.
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.
Berkeley, S. L., Zittel, L. L., Pitney, L. V., & Nichols, S. E. (2001). Locomotor and object control skills of children diagnosed with autism. *Adapted Physical Activity Quarterly, 18*, 405–416.
Bruininks, R., & Bruininks, B. (2005). *Bruininks–Oseretsky test of motor proficiency – 2nd edition manual*. Minneapolis, MN: NCS Pearson.
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the matrix. *Psychological Bulletin, 56*, 81–105.
Chen, W., Faraone, S., Biederman, J., & Tsuang, M. (1994). Diagnostic accuracy of the Children Behavior Checklist scalesfor attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62*, 1017–1025.
Corrigan, B. (2003). Attention deficit hyperactivity disorder in sport: A review. *International Journal of Sports Medicine, 24*, 535–540.
Dodge, K. (1993). Social-cognitive mechanisms in the development of conduct disorder and depression. *Annual Review of Psychology, 44*, 559–584.
De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*, 483–509.
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 Psychology37*.
Drotar, D., Stein, R. E. K., & Perrin, E. C. (1995). Methodological issues in using the child behavior checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology. Special Issue: Methodological Issues in Clinical Child Psychology Research, 24*, 184–192.
DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD rating scale, norms, and clinical interpretation*. New York, NY: Guilford Press.
Efstratopoulou, M., Janssen, R., & Simons, J. (2012). Differentiating children with Attention-Deficit/Hyperactivity Disorder, Conduct Disorder, Learning Disabilities and Autistic Spectrum Disorders by means of their motor behavior characteristics. *Research in Developmental Disabilities, 33*, 196–204.
Egye, C. J., & Short, R. J. (2006). Teacher Self-Efficacy, Burnout, Experience and Decision to Refer a Disruptive Student. *School Psychology International, 27*, 462–474.
Eid, M. (2000). A multitrait–multimethod model with minimal assumptions. *Psychometrika, 65*, 241–261.
Eid, M., Lischetzke, T., & Nussbeck, F. W. (2006). Structural equation models for multitrait–multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: American Psychological Association.
Eid, M., & Diener, E. (Eds.). (2006). *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association.
Epkins, C. C. (1995). Teachers' ratings of inpatient children's depression, anxiety, and aggression: A preliminary comparison between inpatient-facility and community-based teachers' ratings and their correspondence with children's self-reports. *Journal of Clinical Child Psychology, 24*, 63–70.
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*, 141–148.
Henderson, S. E., & Sugden, D. A. (2007). *Movement Assessment Battery for Children. Second Edition*. London: The Psychological Corporation.
Horner, R., Carr, E., Strain, P., Todd, A., & Reed, H. (2002). Problem behavior interventions for young children. *Journal of Autism and Developmental Disorders, 32*, 423–446.
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 and youth* (8th ed.). Columbus, OH: Merrill.
Kazdin, A. E., & Kagan, J. (1994). Models of dysfunction in developmental psychopathology. *Clinical Psychology: Science and Practice, 1*, 35–52.
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.
Kwok, O., Haine, R. A., Sandler, I. N., Ayers, T. S., Wolchik, S. A., & Tein, J. Y. (2005). Positive parenting as a mediator of the relations between parental psychological distress and mental health problems of parentally bereaved children. *Journal of Clinical Child and Adolescent Psychology, 34*, 260–271.
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.
Majdandzic, M., & van den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality, 75*, 121–167.
Marsh, H. W. (1989). Confirmatory factor analysis of multitrait–multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*, 335–361.
Marsh, H. W., & Grayson, D. A. (1995). Latent variable models of multitrait–multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousands Oaks, CA: Sage.
McDermott, P. A. (1994). *National profiles in youth psychopathology: Manual of Adjustment Scales for Children and Adolescents*. Philadelphia: Edumetric and Clinical Science.
Moffit, T. E. (2003). Life-course persistent and adolescence-limited antisocial behavior. In B. B. Lahey, T. E. Moffit, & A. Caspi (Eds.), *Causes of conduct disorder juvenile delinquency* (pp. 49–75). New York: Guilford.
Moffit, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin longitudinal study*. New York: Cambridge University Press.
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*, 1029–1038.
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.
Neuman, R. J., Heath, A., Reich, W., Bucholz, K. K., Madden, P. A. F., Sun, L., et al. (2001). Latent class analysis of ADHD and comorbid symptoms in a population sample of adolescent female twins. *Journal of Child Psychology and Psychiatry, 42*, 933–942.
Newcorn, J. H., Halperin, J. M., Jensen, P. S., Abikoff, H. B., Arnold, L. E., Cantwell, D. P., et al. (2001). Symptom profiles in children with ADHD: Effects of comorbidity and gender. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 137–146.
Ostrov, J. M., & Crick, N. R. (2007). Forms and functions of aggression during early childhood: A short-term longitudinal study. *School Psychology Review, 36*, 22–43.
Piek, J. P., & Dyck, M. J. (2004). Sensory-motor deficits in children with developmental coordination disorder, attention deficit hyperactivity disorder and autistic disorder. *Human Movement Sciences, 23*, 475–488.
Reid, H. M., & Norvilitis, J. M. (2000). Evidence for anomalous lateralization across domains in ADHD children as well as adults identified with the Wender Utah rating scale. *Journal of Psychiatric Research, 34*, 311–316.
Roussos, A., Francis, K., Zoubou, V., Kiprianos, S., Prokopiou, A., & Richardson, C. (2001). The standardization of Achenbach's Youth Self-Report in Greece in a national sample of high school students. *European Child & Adolescent Psychiatry, 10*, 47–53.
Smith, A. L. (2003). Peer relationships in physical activity contexts: A road less traveled in youth sport and exercise psychology research. *Journal of Sport & Exercise Psychology, 4*, 25–39.
SPSS 15.0. (2006). *Statistical package for the social sciences. Command syntax reference*. Chicago, IL: SPSS Inc.

Author's personal copy

Tram, J. M., & Cole, D. A. (2006). A multimethod examination of the stability of depressive symptoms in childhood and adolescence. *Journal of Abnormal Psychology, 115*, 674–686.

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

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.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 68*, 1038–1050.