Quantifying peer interactions for research and clinical use: The Manchester Inventory for Playground Observation

Jenny Gibson a, Jamilla Hussain b, Samina Holsgrove c, Catherine Adams a, Jonathan Green d,

a Dept Human Communication and Deafness, Ellen Wilkinson Building, University of Manchester, Oxford Road, Manchester M13 9PL, UK
b Leeds Teaching Hospitals NHS Trust, Leeds General Infirmary, Great George Street, Leeds LS1 3EX, UK
c Central Manchester University Hospitals Foundation Trust, Trust Headquarters, Cobbett House, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL, UK
d Community Based Medicine, Jean McFarlane Building, University of Manchester, Oxford Road, Manchester M13 9PL, UK

A B S T R A C T

Direct observation of peer relating is potentially a sensitive and ecologically valid measure of child social functioning, but there has been a lack of standardised methods. The Manchester Inventory for Playground Observation (MIPO) was developed as a practical yet rigorous assessment of this kind for 5–11 year olds. We report on the initial reliability and validity of the MIPO and its ability to distinguish social impairments within different psychopathologies.

We observed 144 clinically referred children aged 5:00–11:11 (mean 8.8) years with Externalising (n = 44), Internalising (n = 19), Autism Spectrum Disorders (n = 39) or Specific Language Impairment (n = 42), and 44 class-controls, in naturalistic playground interaction. Observers, blind to clinical diagnosis, completed the MIPO and the teacher checklist from the Social Skills Rating System (SSRS).

MIPO items showed high internal consistency (alpha = .924; all ‘alpha if item deleted’ values > .91), inter-observer reliability (mean $k_w = .77$) and test–retest stability (over 2 weeks; mean $k_w = .58$). MIPO totals showed convergence with SSRS (n = 68, $r_s = .78$, $p < .01$) and excellent discrimination between case and control (sensitivity = .75 and specificity = .88, AUC = .897). Externalising, Autistic Spectrum and Language Impaired groups showed distinct profiles of MIPO impairment consistent with theory: Internalising disorders less so. 65.3% of clinical cases were classified accurately for primary diagnosis.

The MIPO shows reliability and validity as a measure of children’s social functioning relevant in developmental research and as a clinical tool to aid differential diagnosis and intervention planning.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Social difficulties in developmental disorder and psychopathology

Most forms of developmental disorder or child psychopathology are associated with some form of difficulty in social functioning (Guralnick, Hammond, Connor, & Neville, 2006; Hartup, 2005). The pattern of these difficulties will vary with
disorder-type, but are arguably often core to pathogenesis rather than just a functional outcome. Such recognition is hardly new (Sullivan, 1953) but contemporary theory has made the more concrete conceptualisation of ‘social endophenotypes’, intermediate in the pathway to illness expression and a potentially useful marker for illness or proxy target for intervention (Gottesman & Gould, 2003; Skuse et al., 2009). Endophenotypes were originally defined as having an intrinsic or heritable origin (Gottesman & Gould, 2003) but in the context of this paper we propose a broader conceptualisation, which encompasses the variety of social impairment phenotypes seen across childhood development and psychopathology, themselves reflecting different aetiological mechanisms and theories of socialisation. Recent commentary (Banaschewski, 2010) has discussed the implications of this complex variety of social behaviour problems across disorders for multi-level approaches to psychopathology. As examples, we outline below how some different theoretical models of socialisation might predict observed peer social behaviour in children with different disorders and developmental disabilities.

Behaviour genetic studies show that the distributed prosocial and reciprocal social behaviours in the population show significant heritability as well as environmental influence (Plomin, Defries, & McClearn, 1997). In the highly heritable condition of autism, social impairments are a core characteristic of the disorder (Agam, Baulminger, & Shulman, 2003; Baron-Cohen, 2004; Constantino et al., 2004; Holmes & Willoughby, 2005; Hoekstra, Ronald, Happe, Baron-Cohen, & Plomin, 2008), with a wide range of problems in reciprocity. Observed peer behaviours in autism will in theory include aloofness and lack of engagement with others, restricted patterns of play, atypical behaviours and self-stimulation. Social impairments are also seen as functional or co-occurring difficulties within a wide range of complex neuro-development disabilities such as Tourette syndrome, ADHD, specific language impairment (SLI) and developmental coordination disorder. In this wider group the social impairment may be intrinsic or a secondary consequence of primary pathology (for instance in SLI, children’s interpersonal interactive functioning may be compromised by their difficulties in using higher level language to resolve conflict and assert their interests, Horowitz, Jansson, Ljungberg, & Hedenbro, 2005).

Attachment theory emphasises the motivational and goal orientated aspects of prosocial behaviour and plots how certain patterns of early interaction with caregivers impact the child’s social development and response to future social situations (Green, 2009). Primary insecure or disorganised early attachments are associated with cognitive and behavioural social impairments associated with psychopathology (Green & Goldwyn, 2002; Futh, O’Connor, Matias, Green, & Scott, 2008) as well as peer rejection (Futh et al., 2008). Here, theory would imply a set of observed peer behaviours around the presence or absence of general pro-social behaviours, and specific difficulties in care-seeking and care-giving, friendship and intimacy. Children with disorganised insecure attachments often develop conduct disorders and their peer interactions may be rather similar to others in that group. In contrast the more severe patterns of Attachment Disorder are characterised by pervasive social impairments and disinhibition—with core difficulties in social awareness, social appropriateness and reciprocal social functioning that can present at the extreme like autism spectrum disorder (O’Connor & Rutter, 2000). Observed peer social behaviour here will be rather different to conduct problems; with interpersonal awkwardness (for instance indiscriminate friendliness) rather than instrumental aggression.

Social learning theory has been key in plotting the pathways into antisocial behaviour via patterns of escalating coercive and hostile interactive behaviours between parents and children. Children who develop conduct disorders commonly have a bias towards reactivity and hostile attributions to others’ actions, which leads them into conflicted and aggressive social behaviour. Their tendency to lack of inhibitory control and their lack of alternative non-aggressive strategies compound the difficulties. Observed peer interactions would in theory reflect these difficulties by being marked by high frequency of conflict or aggression, difficulty with conflict resolution or bullying, although these children may be relatively skilled in joining or sustaining group activity (Lansford, Malone, Dodge, Pettit, & Bates, 2010). By contrast, the group of children with ‘unsocialised’ conduct disorder would show a different pattern of peer interaction since they are usually more socially isolated with fewer opportunities for reciprocal play (Goodman & Scott, 2005). Children with severe conduct problems in the context of callous and unemotional traits may have a different socialisation pathway again (Frick, Cornell, Barry, Boden, & Dane, 2003; Jones, Happe, Gilbert, Burnett, & Viding, 2010); under-reactive and lacking in social inhibition. Little work has been done on observed peer interaction in this latter group, but theory would suggest observed episodes of cruelty or manipulation of others. Children with internalising disorders may also show social difficulties as a consequence of social withdrawal or social inhibition (Mesman, Bongers, & Koot, 2001). When this becomes pathological the child’s lack of self-esteem and confidence may cause behavioural inhibition and avoidance of new challenges or fear engendering situations; affecting social interaction. Shyness is likely to be seen in observed difficulties in joining, sustaining and involvement in reciprocal play. These children are often seen as being aloof and vulnerable to bullying. They may spend more time alongside teachers and assistants (Goodman & Scott, 2005).

1.2. Naturalistic observation

The presence of such a range of discrete social difficulties across psychopathology and disability implies an equivalent importance for the assessment of social functioning in clinical diagnosis and research (Banaschewski, 2010). Accurate assessments of social impairments may help to characterise disorders and differentiate their phenotypes; they may help understand routes into psychopathology and targets for treatment. Within such assessment, the naturalistic observation of social functioning, and in particular the structured observation of peer interaction, has a number of strengths (Pellegrini, 2001). Specific peer relationship difficulties have been identified across the range of socio-emotional, behavioural and
Communication disorders (Agam et al., 2003; Hartup, 2005; Guralnick et al., 2006). The free-for-all of the playground is a challenging social environment; children engaging in successful peer interaction there demonstrate social-cognitive and communication skills at a more sophisticated level than those evident in other school contexts (Pellegrini, 1998) and nascent social difficulties often first become apparent in that context. Furthermore, the social interaction observed in the playground can reflect key ‘real time’ social information processing that may differ in its mechanism from the more evaluative reasoning assessed in lab-based psychometric tests (Crick & Dodge, 1994). Observation of peer interaction in naturalistic settings is thus likely to be a particularly sensitive assessment context. It is a common clinical observation, for example, that children with high functioning autism can succeed in formal social understanding tasks in the structured environment of the clinic while showing major difficulties in social relating with peers in the community. Finally, observations in the school playground have ecological validity and are in a readymade, reasonably predictable, structured environment for observation. They can complement teacher and parent reports by avoiding reporting bias.

1.3. The Manchester Inventory for Playground Observation

The Manchester Inventory for Playground Observation (MIPO) was designed with these considerations in mind; and in the tradition of systematic observational techniques combining the benefits of naturalistic methods with a commitment to generating valid, objective and reliable data (Salvia & Ysseldyke, 2004). The MIPO aims to be a rigorous, robust yet pragmatically usable observational schedule for social interaction on the playground, useful for both general clinical assessment and research purposes. To this end it was theoretically structured to enable the identification of different patterns of social impairment across disorder, on the assumption that the social impairment associated with different aetiological constructs and psychopathological entities will be manifest within different domains of peer social functioning. Accordingly the scale was constructed in four subscales, which reflect important aspects of peer socialisation for a range of developmental disorders (further details of individual items are outlined in the online Appendix A): A Pro-social Skills subscale codes for those higher-level social skills which children need to facilitate successful interaction. The choice of items has been influenced by developmental theories of peer interaction, which emphasize approaching, and successfully joining an interaction as important skills in peer relationships. A Conflict Management subscale codes the extent of peer-conflict in which the child is involved, as well as the degree to which the child is able to successfully resolve such situations. Peer conflict is not uncommon on the primary school playground but is expected to be exaggerated in externalising disorders where there may be amplified responses to minor conflicts due to impulsivity or attributinal bias. A Confiding and Care-Seeking/Giving subscale reflects the distinction developmental theorists have drawn between peer interactions and friendships: The former referring to skills in joining and sustaining interactions as described in the Pro-social skills subscale while the latter is characterised by shared enjoyment of intimate reciprocal social exchange within a stable dyad (Howes, Rubin, Ross, & French, 1988). The subscale codes for the child’s ability to appropriately give and receive care from peers (and adults who may be present in the playground) and to enjoy engagement in intimate confiding relationships. An Atypicality subscale codes for the presence of unusual behaviours particularly characteristic of Autistic Spectrum and related disorders, which directly impact on the child’s ability to engage in successful peer interaction.

This current study reports on the reliability and validity of this observational method in the context of school playground observations. Two levels of discriminant validity test were undertaken: the first whether the instrument, despite its relatively short observational sampling period, could identify meaningful social impairment in clinical cases compared to age-matched class controls. Secondly, whether it could discriminate different profiles of social impairment between clinical diagnostic groupings along the lines predicted by theory. On the assumption that the instrument should be measuring trait impairment over at least the short term, we also examined test–retest stability to indicate whether variation in playground and activity type over time would affect the measure of the core construct. Criterion validity was examined against an existing instrument measuring a similar construct.

2. Method

2.1. Participants

Participants were recruited from Child and Adolescent Mental Health Services (CAMHS) and Speech and Language Therapy (SLT) services in Greater Manchester, UK. Four Child Psychiatrists within two CAMHS teams made sequential unselected referrals of children with formal clinical diagnoses in groupings of Externalising Disorder (including Conduct Disorder, Oppositional Defiant Disorder, and ADHD), Internalising Disorder (including primary Depression and Anxiety disorders) and Autistic Spectrum Disorder (ASD). In assigning these diagnoses, the referring clinicians used shared diagnostic protocols, standardised assessments and clinical interview. In line with standard UK practice, diagnoses were expressed as a hierarchy, and the primary diagnosis used for allocation in the study with comorbidities noted. Children diagnosed with Specific Language Impairment (SLI) using standard criteria (nonverbal IQ within normal limits, expressive or receptive language at least 1.5 s.d. below mean for age, no sensory disability and no emotional/behavioural difficulties) were recruited from referrals from SLT services.
2.2. Procedures

All children observed were in mainstream school placements and observations took place during the regular playground sessions in the school. Observations were made according to a standard operating procedure and the MIPO scoring protocol. Researchers observed unobtrusively from the edge of the play area but close enough to be able to observe subtle interactions and overhear conversation. The identified referred child was observed naturalistically for 10 min. When school protocols permitted and using procedures derived from other observational instruments (Achenbach & Rescorla, 2001) the researcher identified at random and then observed a class-control child during the same playground session who was engaged in similar play activity to the target child and matched for gender and ethnicity and classroom (age) group. This control child remained anonymous with no further background information obtained, but to minimise the chance of any psychopathology, each control child was rated separately on the ASEBA observational schedule for school-age children (Achenbach & Rescorla, 2001)—no psychopathology was found. Directly after the observation period, a detailed narrative summary of the observations was written and used as an aide-memoire to assist the immediate rating of both identified and control children on MIPO items.

The three research observers were postgraduate students from medicine, speech and language therapy and psychology. They were trained on all instruments and undertook up to eight supervised MIPO observations to ensure training reliability.

The study design was approved by the Wrightington, Wigan & Leigh Local Research Ethics Committee (07/H1014/70). Research observers obtained written informed consent from parents of referred clinical cases.

2.2.1. Measures

The MIPO consists of 28 items arranged in the 4 subscales described above. Items within each subscale have a descriptive definition and are scored according to a four category, rank-ordered rating scale:

(1) The child is socially competent within this domain.
(2) Evidence of sporadic or minimal difficulties with an item but with no significant functional consequence.
(3) Difficulty in a particular area of moderate intensity and functional consequence.
(4) Marked difficulties in an area, either in intensity or frequency of recurrence. No evidence of competency and one or more instances in which the intensity and/or functional consequences are significantly socially inappropriate or damaging.

Rating is assisted through a series of descriptive behavioural anchors provided for each item. If no opportunity to observe a given behaviour arose, the coder assigned an arbitrary score of 8, treated as zero for the purpose of analysis. All subscales are scored so that a higher score indicates a greater degree of impairment.

To evidence construct validity, (Kane, 2001) scores from the experimental scale (MIPO) were compared with an established measure social functioning, the Social Skills Rating System (SRSS, (Gresham & Elliott, 1990)). SRSS is a checklist-based measure of social functioning with separate scales for completion by parents, teachers and children. For the purpose of the current study we used teacher completed versions of SRSS as we felt teacher-reported school-based social competence most closely reflected the aims of MIPO.

2.2.2. Measures to minimise bias

The researchers were unknown to the children at the time of their visit and steps taken so that neither the proband nor control child was aware that they specifically were being observed. Independent staff undertook all administration of referrals and allocation to researchers. Research observations of the referred child were made blind to their specific diagnosis and observations for inter-observer reliability were summarised and scored independently by each observer.

2.3. Planned data analyses

2.3.1. Reliability

Internal consistency of the full scale and the subscales of MIPO was calculated using Cronbach’s alpha. Inter-observer reliability was undertaken using independent simultaneous ratings by two observers who independently scored their observations according to MIPO criteria. Scores were compared at item level using Cohen’s kappa and weighted kappa coefficient (κ, κw). Kappa was considered to be the most appropriate coefficient as the comparisons made were between categorical ratings. κ gives an indication of the extent to which observers rated the child in the same likert category, whereas κw reflects how close ratings were, for example ratings in adjacent categories (e.g., 1 and 2) result in a higher κw than ratings in non-adjacent categories (e.g., 1 and 4). Test–retest observations over 7–14 days were carried out using the same observer and, as far as possible, at the same playtime and day of week. Time 1 and Time 2 scores for each item were compared using Cohen’s kappa coefficient (κ).

2.3.2. Validity

Discriminant validity of the MIPO in distinguishing clinical cases from controls was tested using a receiver–operator characteristic (ROC) curve, a plot of the sensitivity against 1 minus the specificity for each potential cut-off score on MIPO.
The area under curve (AUC) was calculated as a summary of MIPO’s discriminative power at this level. Discrimination between the different clinical groups was investigated using discriminant function analysis (DFA). DFA uses data from continuously distributed variables (in this case, the MIPO subscales) to predict category membership (diagnostic group), the accuracy of the predicted classifications therefore tells us something about the utility of MIPO in diagnostic decisions and the characteristics of peer social functioning impairments typical of each group.

Convergent validity between MIPO and SSRS was tested using Spearman’s correlation coefficient ($r_s$) for non-parametric data (Howell, 1997).

3. Results

3.1. Sample

Descriptive characteristics of the clinical sample are presented in Table 1. In 43% of the sample there was one comorbid diagnosis and in 5.6% there were two or more comorbid diagnoses. Mean age was 8.8 years; ethnicity included 88% white British, 5.9%, Asian British, 5.9%, black British. The different clinical diagnostic groupings did not differ significantly in distribution of age or gender. Presence of co-occurring diagnoses was associated with more impaired scores on all sub-scales except that for Conflict (Pro-social, $t(103) = -3.09$, $p < .001$, $r = .3$; Conflict, $t(103) = -0.29$, ns; Confiding, $t(103) = -3.80$, $p < .001$, $r = .4$; Atypicality, $t(103) = -3.02$, $p < .001$, $r = .3$).

3.2. Reliability

Cronbach’s alpha ($\alpha$) for the whole scale was .924; an alpha if item deleted analysis showed that no item if deleted would reduce this alpha to a value below .9, suggesting that the consistency of the scale was not dependent on a few items. For subscales, values of $\alpha$ were: Pro-Social $\alpha = .885$, Conflict $\alpha = .834$, Care/Confiding $\alpha = .723$, and Atypicality $\alpha = .751$. Estimates of inter-observer reliability and test–retest stability of MIPO items were made from random subsamples of 27% (39/144) and 14% (20/144) of the clinical cases respectively, and are set out in Table 2.

<table>
<thead>
<tr>
<th>Diagnostic category</th>
<th>n</th>
<th>Mean age (SD)</th>
<th>Male:female</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>39</td>
<td>8.2 (1.7)</td>
<td>32:7</td>
</tr>
<tr>
<td>Externalising disorder</td>
<td>44</td>
<td>9.5 (1.5)</td>
<td>41:3</td>
</tr>
<tr>
<td>Internalising disorder</td>
<td>19</td>
<td>9.5 (2.1)</td>
<td>14:5</td>
</tr>
<tr>
<td>SLI</td>
<td>42</td>
<td>8.4 (1.6)</td>
<td>34:8</td>
</tr>
<tr>
<td>Total clinical sample</td>
<td>144</td>
<td>8.8 (1.7)</td>
<td>121:23</td>
</tr>
</tbody>
</table>

Fig. 1. ROC curve for cases and controls classified by MIPO.
3.3. Validity

Sensitivity and specificity of MIPO scores in discriminating case vs controls was studied using all clinical cases compared to data on the 44 class-controls (approval was not given to collect control data in all cases). Fig. 1 shows the receiver-operator characteristic (ROC) curve for this analysis; optimum MIPO cut-off score for discriminating between case and control was 13, where sensitivity = 0.75 and specificity = 0.88. The Area under the Curve was 0.897.

The distinct profiles of MIPO subscale means across separate clinical diagnostic groups are shown in Fig. 2. Discriminant function analysis within the clinical sample was performed on logarithmically transformed data since it did not meet assumptions of multivariate normality. Analysis revealed two significant functions, the first accounting for 57.6% (Wilks’ lambda = .367, χ² = 139.20, df 12, p < .001) and the second 42.1% (Wilks’ lambda = .643, χ² = 61.38, df 6, p < .001) of the variance. A 3rd function accounted for the remaining 0.3% of the variance but was excluded from the analysis as it was non-significant. A structure matrix of correlations between the subscale scores and standardised discriminant functions is presented in Table 3. Classification accuracy was assessed after adjustment of prior probabilities for group size. Successful classification rates were: ASD, 69.2%, Externalising 75%, SLI 81%, Internalising 0%. In total, 65.3% of cases from the clinical sample were correctly classified.

Analysis of convergent validity of MIPO with teacher SSRS scores was undertaken on the subset of 68/144 children for which SSRS data was available. This subset did not differ from the whole clinical group on any variable except in a marginally higher total MIPO score (F = 4.013, p = 0.047). SSRS scores were significantly correlated with total MIPO scores, r = .782, p < .001, one-tailed.

### Table 2
Mean values of κ and κw for inter-observer and test-retest comparisons.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Inter-observer κ (range)</th>
<th>κw (range)</th>
<th>Test-retest κ (range)</th>
<th>κw (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Pro-social</td>
<td>.706 (.414–.958)</td>
<td>.790 (.612–.981)</td>
<td>.487 (.167–.695)</td>
<td>.560 (.286–.856)</td>
</tr>
<tr>
<td>B: Conflict</td>
<td>.678 (.516–.821)</td>
<td>.790 (.473–.893)</td>
<td>.511 (.355–.674)</td>
<td>.600 (.529–.673)</td>
</tr>
<tr>
<td>C: Confiding</td>
<td>.631 (.612–.753)</td>
<td>.698 (.675–.841)</td>
<td>.317 (.185–.457)</td>
<td>.481 (.052–.647)</td>
</tr>
<tr>
<td>D: Atypicality</td>
<td>.828 (.629–1)</td>
<td>.850 (.635–1)</td>
<td>.527 (.251–.770)</td>
<td>.616 (.440–.809)</td>
</tr>
<tr>
<td>Full scale</td>
<td>.703</td>
<td>.774</td>
<td>.488</td>
<td>.577</td>
</tr>
</tbody>
</table>

### Fig. 2
MIPO subscale profiles by diagnosis.

### Table 3
Correlations between predictor variables and discriminant functions.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C: Confiding</td>
<td>.909</td>
<td>.150</td>
</tr>
<tr>
<td>B: Conflict</td>
<td>.701</td>
<td>-.248</td>
</tr>
<tr>
<td>A: Pro-social</td>
<td>.632</td>
<td>.570</td>
</tr>
<tr>
<td>D: Atypicality</td>
<td>.059</td>
<td>.971</td>
</tr>
</tbody>
</table>
4. Discussion

Inter-observer reliability on the MIPO was good, with the mean values for the scale falling within the “substantial agreement” range (Landis & Koch, 1977). Eight of the individual MIPO items – ‘Joining’, ‘Sustaining Social Interaction’, ‘Social Reciprocity’, ‘Conflict Intensity’, ‘Conflict instigator’, ‘Atypicality’, ‘Aloneness’ and ‘Verbal Interaction’ – had κ greater than or equal to .8, falling within Landis and Koch’s “almost perfect agreement” category (Landis & Koch, 1977). No item had a kappa of a value less than .41, the lower bound for “moderate agreement” (Landis & Koch, 1977) but two items – ‘Follows Group Rules’ (κ = .414) and ‘Gives Care’ (κ = .414) – approached this. The former had a weighted κ of .637 indicating that disagreements were largely in adjacent scoring categories and suggesting that its behavioural descriptors might be underspecified and need amplification in future studies. Weighted kappa for ‘Gives Care’ was a still moderate .487; possibly reflecting the fact that this behaviour was only codable in the context of care for another and was relatively frequently coded 8.

The Cronbach’s alpha value of .944 supports the conclusion that MIPO is consistently measuring an underlying behavioural construct. Furthermore, the high level of convergence with the SRSS (Spearman’s r = .82) suggests that the construct MIPO quantifies is valid as a measure of social competence with peers. Further evidence of validity comes from the highly sensitive and specific discrimination of caseness from control, summarised by the area under ROC curve. However, this result should be interpreted with caution as observers were not blind to the status of control children and observer bias cannot be ruled out.

The test–retest analysis yielded mean agreement values falling within the “moderate agreement” range (κ = .468, \( \kappa_{rs} = .577 \), for the full scale). Some items showed greater stability than others, notably ‘Joining’, ‘Social Reciprocity’, ‘Friendship Group’, ‘Enjoyment of Play’, ‘Follows Group Rules’, ‘Atypicality’ and ‘Verbal Interaction’, all of which had \( \kappa_{rs} > .7 \). A few items – ‘Friends with Younger Children/Adults’, ‘Responds well to a Good Natured Tease’, ‘Sticks up for Self’ and ‘Responds to Care’ – had \( \kappa_{rs} < .4 \). Given the number of variables within the context of the playground, it is encouraging that the majority of MIPO items reflect measurement of stable behavioural traits rather than being subject to changing environmental context. While relatively unstable items could be removed from the scale, we consider a preferable option at this stage would be a further study using generalisability (g) theory (Dunn, 1999; Lei, Smith, & Suen, 2007) to further identify potential sources of non-random variation in the results.

The results from the discriminant analysis suggest a number of interesting clinical and theoretical implications. Examined of the structure matrix (Table 3) suggests that the Confiding and Conflict subscales contributed most to Function 1, while Atypicality contributed most to Function 2. For Function 2 higher levels of Atypicality were negatively correlated with Conflict scores, possibly reflecting the lower levels of social engagement of any kind for those with elevated autistic symptomatology. The Pro-Social subscale was important for both functions, indicating the high level of pro-social impairment present in all of the diagnostic groups in this study. Externalising, SLI and ASD groups showed distinct profiles of impairment consistent with theory (Fig. 2); for example, the SLI group showed consistently lower levels of social impairment than children with psychiatric conditions, the Externalising group showed elevated levels of conflict and the ASD group elevated levels of atypical behaviours. There was poorer discrimination for Internalising disorders. These showed overall raised levels of social difficulty against controls but a relatively even profile of difficulty across subscales and a lack of correct classification on the 2 functions. This poorer discrimination may partly be due to the much smaller sample size for this group (itself reflecting the lower rates of identification of primary internalising disorder in clinical practice). Also of interest are some observed misclassifications. All children in the SLI group had by definition scored below ASD threshold on independent ADOS, but all the misclassifications in this group were with the ASD category. Conti-Ramsden, Simkin, and Botting (2006) found in a longitudinal study of language-impaired children that a subgroup of SLI children went on to meet criteria for ASD in their adolescent years and hypothesised aetiological continuity between the two conditions. It could be that MIPO scores here are picking up the first signs of a subtle shift towards autism-like social deficits in the social functioning of children with language impairments.

This initial data therefore suggests that it is possible, using this new instrument, to make systematic observations of peer interactions within a playground setting that yield valid data about a child’s peer interaction skills. There is evidence that that distinct patterns of social impairment associated with different psychiatric disorders in childhood can be identified in this way. The main limitation of the study is the relatively small diagnostic group sizes in order to allow for the inevitable heterogeneity within different clinical groups, and larger-scale studies will be necessary to confirm and explore further these findings. Another relative limitation for analysis stemmed from the lack of permission to code control children in some sites and difficulty getting SSRS returns. There is ethical sensitivity about videotaping peer groups in naturalistic settings and the MIPO therefore is constrained in our experience to be a live-coded measure in most schools in the UK. This did have the benefit of making the observations less obtrusive (an important advantage), but the limitation that it was not possible to make blinded codings in relation to case vs control children (although coders were blind to diagnostic group).

Direct observational methods such as MIPO have high ecological validity but have been criticised for being highly subjective and producing data which are not readily generalisable beyond the immediate time and situation of a specific observational episode (Pellegrini, Symmons, & Hoch, 2004). MIPO has been developed to address these issues through use of clearly operationalised definitions of the behaviours of interest, adequate training for observers, standardised scoring procedures and sound theoretical rationale. Reliability and validity are typically hard to establish for naturalistic observational methods (Hintze, 2005) however the results presented in this paper are encouraging in that at least moderately reliable data can be gained from a 10 min observation.
5. Conclusion

The MIPO playground observation is practical to carry out in a typical school assessment visit and provides vivid data on real-time social functioning with peers. Because we have shown the short-term stability of a MIPO observation, the results could be considered indicative within an assessment, and provide a cost-effective complement to other domains of assessment such as history-taking, clinical examination, psychometric examination and laboratory based social cognition tasks, all of which can be time and staff intensive. The data suggests that the examination of sub-scale profiles could give discriminating information on social functioning to help differential diagnosis—for instance as between conduct and subtle autistic spectrum disorders. It could also help to clarify social deficits in naturalistic settings as part of a functional assessment across disorders. The specificity and discriminating nature of this information contrasts with the commonly non-systematic and generalised information about a child’s peer functioning which is currently gathered from other sources in clinical practice. Typically, clinical assessments of peer functioning rely on indirect parental and teacher reports; while these indirect methods are important, they have limitations: teacher ratings are subject to bias from halo effects and often correlate poorly with observational data regarding peer interaction (Abikoff, Courtney, Pelham, & Koplewicz, 1993).

Information such as that from the MIPO has a further potential role within treatment planning, helping the profiling of social impairments as a useful intermediate target for intervention instead of disorder-specific outcome behaviours. For example children scoring high on the “Joining” item could be explicitly taught strategies for joining peer interactions using comic strip conversations, role-play and structured practise. A number of social skills interventions could be conceptualised and targeted in this way. Such a strategy is particularly relevant in the social domain, since the nature of social relationships is to be self-reinforcing and difficulties in social interaction if untreated are likely to become maintaining or amplifying factors for psychopathology in development.

In research, the MIPO holds promise to be a useful tool in the investigation of functional social impairment in different disorders. Its structure, based on an implicit theoretical model of social impairment discrete to the aetiology of specific disorders, could help avoid errors consequent on applying measures developed for one disorder onto another. In contrast, it could also illuminate the extent to which common endophenotypes may act across disorders, and help with clarification of developmental processes. Observational methods of this kind in naturalistic settings can combine with peer nomination, individual clinical assessment and different respondent reports to provide a powerful triangulation with which to identify varieties of social impairment in development.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ridd.2011.07.014.

References


