Convergence and concurrent validity of DSM-III diagnoses and the Personality Inventory for Children (PIC)

REX B. KLINE
Concordia University

DAVID LACHAR
University of Texas Medical School at Houston

AND

CHARLES L. GDOWSKI
Lafayette Clinic and Wayne State University

ABSTRACT

Psychiatric diagnoses and objective behaviour rating scales represent alternative means to describe children's psychological disorders. Results of several studies regarding the convergence of diagnoses and rating scales have been mixed, raising questions about the nature of the relation between the two systems. The purpose of this study was to evaluate the relations of DSM-III diagnoses and the parent-informant Personality Inventory for Children (PIC) within a sample of behaviourally disturbed children and adolescents. In addition, the ability of DSM-III diagnoses and the PIC to predict teacher and clinician behaviour ratings was compared. High proportions of children assigned specific DSM-III diagnoses obtained elevated scores on related PIC scales, though the overlap was imperfect. Also, covariation of the teacher and clinician behaviour ratings was much stronger with the PIC than with DSM-III diagnoses. Implications of these findings for clinical practice with children are discussed.

There are two major approaches to the classification of child psychological problems. Psychiatric diagnostic typologies such as the Diagnostic and Statistical Manual (DSM-III, American Psychiatric Association, 1980) represent the most traditional means of describing child mental health status. The major focus of this approach has been the delineation of syndromes (defined by specific symptom clusters), and the delineation of differential diagnostic criteria. A more recent approach has been characterized by statistical identification of dimensions of adjustment (typically using factor analysis), and the construction of objective rating scales to measure each behavioural dimension (Achenbach & Edelbrock, 1978). This approach arose in part because of longstanding concern about low reliability and questionable external validity of diagnostic categories for children (e.g., Mattison, Cantwell, Russell, & Will, 1979; Mezzich, Mezzich, & Coffman, 1985; Rutter & Shaffer, 1980; Werry, Methven, Fitzpatrick, & Dixon, 1983).

Because diagnoses and rating scales are alternative ways to describe children referred for psychological consultation there has been interest in evaluating their...
convergence. Diagnosis-rating scale correspondence has often been assessed using indices of sensitivity and specificity (e.g., Bird, Canino, Gould, Ribera, Rubio-Stipec, Woodbury, Huertas-Goldman, & Sesman, 1987; Kazdin, Colbus, & Rodgers, 1986; Kazdin & Heidish, 1984; Steinhausen & Gobel, 1987). These indices were originally developed in medicine to indicate how accurately screening tests could indicate the presence of a disease in asymptomatic patients (Griner, Mayewski, Mushlin, & Greenlan, 1981). Sensitivity is calculated as the proportion of diseased patients who have a positive result on the test, while specificity is the proportion of control subjects who are not diseased who have negative (normal) outcomes on the test.

Sensitivity is estimated in studies of diagnosis-rating scale convergence as the proportion of children who receive a diagnosis (e.g., Depression) and who also receive clinical-range scores on an objective rating scale assessing the same domain (e.g., \( T > 69 \) on the Depression scale of the Child Behaviour Checklist; CBCL; Achenbach, 1978). A positive psychiatric diagnosis along with a clinical-range score on a behavioural checklist provides evidence of the concurrent validity of both approaches.

Results regarding the sensitivity of parent-informant checklists have been mixed. Kazdin and Heidish (1984), Kazdin et al. (1986), and Lobovits and Handal (1985) reported that the proportions of inpatient and outpatient children assigned diagnoses of Depression who also obtained elevated scores on several parent-informant checklists of child depression ranged from 44 to 100% (average sensitivity across all 3 studies = 73.4%). Bird et al. (1987) evaluated the ability of CBCL scales to identify children who received any DSM-III Axis I or II diagnosis in an epidemiological survey, and these investigators reported sensitivity values ranging from 67% to 72% (average = 66%).

Other researchers have reported much lower sensitivity values for parent-informant rating scales. Shekim et al. (1986) administered the CBCL to mothers of 14 children assigned DSM-III diagnoses of Attention Deficit Disorder with Hyperactivity, and only two children received elevated scores on the CBCL Hyperactive scale (sensitivity = 14.3%). Steinhausen and Gobel (1987) examined the correspondence of ICD-9 diagnoses of Hyperactivity, Conduct Disorder, Emotional Disorder, and Mixed Conduct and Emotional Disorder and the Child Behav-

\[ ^{1} \text{Interrater reliabilities of these diagnoses were not estimated in this study, but estimates (kappa coefficients) for the following categories have been reported by Werry, Methven, Fitzpatrick, and Dixon (1983; WMFD), Mezzich, Mezzich, and Coffman (1985; MMC), Lobovits and Handal (1985; LH), and Bird, Canino, Gould, Ribera, Rubio-Stipec, Woodbury, Huertas-Goldman, and Sesman (1987; BCGRRWHS); Depression (BCGRRWHS: .46); Overanxious Disorder (WMFD: .65); Adjustment Reaction (MMC: .05); Mental Retardation (WMFD: .62; MMC: .96; BCGRRWHS: .85); Conduct Disorder (WMFD: .53; MMC: .43; BCGRRWHS: .31); Developmental Disorder (BCGRRWHS: .31); Schizophrenic Disorder (WMFD: .70); Attention Deficit Disorder (WMFD: .73).} \]
ior Questionnaire (Rutter, Tizard, & Whitmore, 1970), and found low sensitivity values for most scales (range = 14.6% to 42.1%, average = 27.8%).

Specificity is estimated as the proportion of children who have not received a diagnosis (such as Conduct Disorder) who obtain normal-range scores on a screening test for that disorder (such as a delinquency scale). Kline (in press) has identified a conceptual problem, however, when specificity of objective rating scales is estimated in this way. An example from the Kazdin and Heidish (1984) study illustrates this difficulty. In this study 26 of 113 inpatient cases were diagnosed as depressed, and of these a total of 23 obtained an elevated score on the CBCL Depression scale (sensitivity = 23/26 = 88.5%). Among the remaining 87 cases (who received other diagnoses), only 17 obtained normal-range Depression scale scores (specificity = 17/87 = 19.5%).

The meaning of low specificity in medical literature is that the test yields many false positive results. To apply such an interpretation in the above example would require the assumption that none of these 87 conduct disorder children showed clinical symptoms of depression. Absence of a diagnosis of depression in the presence of an elevated CBCL depression scale score would therefore indicate a false positive test result. This is probably an untenable assumption, particularly for children with multiple behaviour problems for whom a single diagnostic label would describe only some of their problems (Kline, in press). Also, while it is possible to assign multiple diagnoses in some diagnostic schemes (such as multiple Axis I diagnoses in DSM-III), there is doubt how often this occurs in clinical assessment (Rutter & Shaffer, 1980) and about the reliability of this practice (Cantwell, Mattison, Russell, & Will, 1979). Without diagnostic information which specifies the full range of problems of children, specificity indices should not be used.

While it is important to evaluate the convergence of diagnoses and rating scales, it is equally crucial to recognize that neither of these sources of information constitute an ultimate criterion or "gold standard" in the evaluation of children. Several other types of information (e.g., objective teacher reports, structured interviews of children) can, and should, be incorporated in clinical assessments. Ideally, diagnoses, rating scales, and other assessment modalities should be used to describe the etiology and treatment of childhood problems. Because of limited knowledge of causes and optimal treatment strategies for children's psychological problems, however, it is probably realistic to expect only that any of these sources of information will provide accurate descriptions of current behavioural status (Zigler & Phillips, 1961).

The purpose of the present study was two-fold. First, the convergence of DSM-III diagnoses and scores from a parent-informant rating scale — the Personality Inventory for Children (PIC; Wirt, Lachar, Klinedinst, & Seat, 1984) — among behaviourally disturbed children and adolescents was evaluated. The proportions of children assigned nine different types of DSM-III diagnoses who also had elevated scores on related PIC scales were determined, and sensitivity values for the
TABLE 1
Demographic characteristics of DSM samples

<table>
<thead>
<tr>
<th>DSM Diagnostic Groups</th>
<th>DP</th>
<th>OA</th>
<th>CD</th>
<th>AH</th>
<th>MR</th>
<th>DD</th>
<th>SX</th>
<th>IA</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>127</td>
<td>96</td>
<td>157</td>
<td>80</td>
<td>47</td>
<td>40</td>
<td>31</td>
<td>34</td>
<td>227</td>
</tr>
<tr>
<td>Age ± (yrs)</td>
<td>11.8</td>
<td>10.8</td>
<td>11.5</td>
<td>9.6</td>
<td>10.0</td>
<td>10.9</td>
<td>12.3</td>
<td>10.2</td>
<td>11.0</td>
</tr>
<tr>
<td>SD</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.4</td>
<td>1.9</td>
<td>3.4</td>
<td>3.3</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Range</td>
<td>4-16</td>
<td>4-16</td>
<td>4-16</td>
<td>3-16</td>
<td>6-14</td>
<td>3-15</td>
<td>8-16</td>
<td>3-16</td>
<td>3-16</td>
</tr>
<tr>
<td>% male</td>
<td>61</td>
<td>69</td>
<td>66</td>
<td>82</td>
<td>75</td>
<td>83</td>
<td>64</td>
<td>83</td>
<td>73</td>
</tr>
<tr>
<td>% White</td>
<td>55</td>
<td>60</td>
<td>47</td>
<td>60</td>
<td>68</td>
<td>52</td>
<td>61</td>
<td>52</td>
<td>51</td>
</tr>
<tr>
<td>% Black</td>
<td>45</td>
<td>40</td>
<td>53</td>
<td>40</td>
<td>40</td>
<td>48</td>
<td>48</td>
<td>49</td>
<td>49</td>
</tr>
</tbody>
</table>

Note: DSM groups: DP = Depression; OA = Overanxious; CD = Conduct Disorder; AH = Attention Deficit Disorder with Hyperactivity; MR = Mental Retardation; DD = Developmental Disorder; SX = Schizophrenia; IA = Infantile Autism; AR = Adjustment Reaction.

PIC were derived. The second purpose was to determine the relative ability of DMS-III diagnoses and the PIC to predict symptom checklist ratings of classroom teachers and interviewing clinicians. While teacher and clinician checklists do not, of course, constitute "gold standards" against which to evaluate the usefulness of diagnoses or the PIC, they can provide convenient, reliable (Achenbach & Edelbrock, 1978), and useful information about current child functioning and can aid in estimating the concurrent validity of these two approaches.

METHOD

Subjects
The data base for this study consisted of child psychiatry evaluations (consisting of mother, teacher, and clinician ratings) of 839 children and adolescents. These cases were evaluated at a large, urban child mental health facility in the midwestern United States, and 70% of the subjects' households fell within the lowest two Hollingshead (1975) socioeconomic categories.

The sample was partitioned into nine groups according to the DSM-III diagnosis each child received, including Depression, Overanxious Disorder, Adjustment Reaction with Mixed Disturbance of Emotions and Conduct, Mental Retardation, Conduct Disorder, Mixed Specific Developmental Disorder (Axis II), Schizophrenic Disorder, Attention Deficit Disorder with Hyperactivity, and Infantile Autism. The demographic characteristics of these samples are reported in Table 1. These nine diagnostic groups differed significantly with regard to sex ($X^2 (8) = 17.06, p < .05$) and age ($F (8, 720) = 4.08, p < .01$) but not race ($X^2 (8) = 9.70, p > .05$), and these differences were consistent with expected characteristics of such diagnostic groups. For example, the cases that received diagnoses of Attention Deficit Disorder were the youngest and had the second highest representation of males. The cases who received diagnoses of Schizophrenia, on the other hand, had the highest mean age.

Measures
PIC. The mothers of each child completed full (600 items) PIC inventories. The PIC is an objective, multidimensional measure of child and adolescent behaviour, emotional, and cognitive status. The standard profile scales, constructed using either empirical-keying or rational/content methods, include three scales that measure informant response set, Lie (L), Frequency (F), and Defensiveness (DEF);
### TABLE 2
Percent of DSM diagnostic groups obtaining clinically elevated scores on PIC scales

<table>
<thead>
<tr>
<th>PIC</th>
<th>DP</th>
<th>OA</th>
<th>CD</th>
<th>AH</th>
<th>MR</th>
<th>DD</th>
<th>SX</th>
<th>IA</th>
<th>AR</th>
<th>Base Rate</th>
<th>X²(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>108</td>
<td>80</td>
<td>132</td>
<td>72</td>
<td>44</td>
<td>34</td>
<td>28</td>
<td>29</td>
<td>202</td>
<td>729</td>
<td></td>
</tr>
<tr>
<td><strong>Internalizing Scales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOM</td>
<td>39</td>
<td>39</td>
<td>33</td>
<td>21</td>
<td>21*</td>
<td>47</td>
<td>36</td>
<td>0*</td>
<td>33</td>
<td>33</td>
<td>28.16**</td>
</tr>
<tr>
<td>D</td>
<td>53</td>
<td>50</td>
<td>50</td>
<td>43</td>
<td>18**</td>
<td>44</td>
<td>54</td>
<td>46</td>
<td>46</td>
<td>48</td>
<td>18.34*</td>
</tr>
<tr>
<td>WDL</td>
<td>38*</td>
<td>25</td>
<td>27</td>
<td>11**</td>
<td>21</td>
<td>27</td>
<td>32</td>
<td>25</td>
<td>25</td>
<td>27</td>
<td>17.83*</td>
</tr>
<tr>
<td>ANX</td>
<td>36</td>
<td>43</td>
<td>30</td>
<td>38</td>
<td>13**</td>
<td>32</td>
<td>29</td>
<td>7**</td>
<td>33</td>
<td>33</td>
<td>22.50**</td>
</tr>
<tr>
<td><strong>Externalizing Scales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLQ</td>
<td>53</td>
<td>45</td>
<td>64*</td>
<td>63</td>
<td>7**</td>
<td>50</td>
<td>61</td>
<td>7**</td>
<td>58</td>
<td>53</td>
<td>22.49**</td>
</tr>
<tr>
<td>HPR</td>
<td>50</td>
<td>53</td>
<td>66</td>
<td>83**</td>
<td>30**</td>
<td>62</td>
<td>43</td>
<td>24**</td>
<td>60</td>
<td>59</td>
<td>56.27**</td>
</tr>
<tr>
<td><strong>Cognitive Scales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACH</td>
<td>61</td>
<td>60</td>
<td>57</td>
<td>67</td>
<td>98**</td>
<td>77</td>
<td>79</td>
<td>76</td>
<td>51**</td>
<td>64</td>
<td>46.12**</td>
</tr>
<tr>
<td>IS</td>
<td>55**</td>
<td>36</td>
<td>28**</td>
<td>38</td>
<td>91**</td>
<td>38</td>
<td>39</td>
<td>79**</td>
<td>28**</td>
<td>42</td>
<td>96.39**</td>
</tr>
<tr>
<td>DVL</td>
<td>51</td>
<td>48</td>
<td>46*</td>
<td>55</td>
<td>93**</td>
<td>65</td>
<td>64</td>
<td>83**</td>
<td>44**</td>
<td>55</td>
<td>52.34**</td>
</tr>
<tr>
<td><strong>Other Scales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAM</td>
<td>54</td>
<td>58</td>
<td>49</td>
<td>53</td>
<td>25**</td>
<td>53</td>
<td>46</td>
<td>45</td>
<td>54</td>
<td>54</td>
<td>22.01**</td>
</tr>
<tr>
<td>PSY</td>
<td>35*</td>
<td>26**</td>
<td>39</td>
<td>37</td>
<td>55</td>
<td>38</td>
<td>61</td>
<td>97**</td>
<td>34**</td>
<td>46</td>
<td>63.45**</td>
</tr>
<tr>
<td>SSK</td>
<td>44</td>
<td>43</td>
<td>46</td>
<td>57</td>
<td>39</td>
<td>32</td>
<td>46</td>
<td>83**</td>
<td>44</td>
<td>46</td>
<td>23.78**</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

Note: DSM groups: DP = Depression; OA = Overanxious; CD = Conduct Disorder; AH = Attention Deficit with Disorder with Hyperactivity; MR = Mental Retardation; DD = Developmental Disorder; SX = Schizophrenia; IA = Infantile Autism; AR = Adjustment Reaction.
a general screening scale for adjustment problems, Adjustment (ADJ); and 12 clinical scales, including three reflecting cognitive functioning, Achievement (ACH), Intellectual Screening (IS), and Development (DVL); two measuring externalizing behaviour, Delinquency (DLQ) and Hyperactivity (HPR); four reflecting internalizing problems, Somatic Concern (SOM), Depression (D), Withdrawal (WDL), and Anxiety (ANX); and others reflecting interpersonal functioning, Psychosis (PSY) and Social Skills (SSK), and family or marital conflict, Family Relations (FAM).

Scores for all PIC scales are expressed in T-score units, and high scores reflect greater psychopathology. The normative sample was comprised of 2582 children with no previous referrals to mental health facilities, and PIC scales are normed separately by sex for ages 3–5 and 6–16; IS is normed for ages 3–5, 6, 7, 8, 9, and 10–16. Test-retest reliability coefficients for PIC scales are satisfactory; Wirt et al. (1984) reported two-week test-retest coefficients ranging from .46 to .94 (average = .86) for an outpatient sample, and two-month test-retest coefficients ranging from .39 to .89 (average = .71) for a normal sample.

Several studies have established the external validity of the PIC with several populations, including cognitively impaired (Clarke, Kehle, Bullock, & Jenson, 1987; Kline, Lachar, & Sprague, 1985), hyperactive (Breen & Barkley, 1983; Forbes, 1985), developmentally impaired (Kline, Maltz, Lachar, Spector, & Fischhoff, 1987), and affective disorder (Lobovits & Handal, 1985; Nieman & Delong, 1987) children. Actuarial interpretive systems have been developed for individual PIC scales (Lachar & Gdowski, 1979; Lachar, 1982) and whole profile configurations (Kline, Lachar, & Gdowski, 1987). Other validity studies are summarized in the test manual (Wirt et al., 1984).

**Teacher and Clinician Ratings.** Classroom teachers and interviewing clinicians completed different behavioural checklists developed by Lachar, Gdowski, & Snyder (1984) for all cases in each diagnostic group except the Mental Retardation and Infantile Autism samples. Each checklist was constructed with factor analysis and consisted of seven factors. The score for an individual child on each factor is simply the number of items (presented in a true-false format) endorsed by the rater. Higher factor scores reflected greater pathology in the areas measured by these dimensions.

The specific item composition of these 14 checklist factors is presented in Lachar et al. (1984). Checklist factors, number of items, internal consistency (alpha) coefficients, and example items are presented below. Teacher dimensions: Hostility/Impulsivity (24/.94/temper tantrums; steals from other children); Poor Study Skills (12/.89/does not pay attention; does not complete assignments); Academic Delay (11/.89/below average reading; difficulty understanding directions); Poor Classroom Adjustment (12/.87/rejects school; poor attendance); Poor Self-Concept/Depressive Symptoms (11/.80/refers to self as dumb; lonely, unhappy); Social Withdrawal (8/.77/has no friends; no group activities); and Distractibility/Motor Restlessness (9/.77/easily distracted; overactive). Clinician dimensions: Hostility/Dyscontrol (14/.87/fights with peers; fire setting); Language/Motor Deficits (11/.79/fine motor ataxia; receptive aphasic symptoms); Emotional Lability/Impulsivity (13/.79/rapid mood shifts; destroys objects); Disorganization/Poor Reality Testing (8/.72/inappropriate affect; thought disorder); Depressive/Somatic Symptoms (11/.66/frequent crying; sadness); Antisocial Behaviour (7/.66/truancy; police involvement); and Social Withdrawal (8/.66/isolative; excessive shyness).

**Content Overlap of PIC and Teacher and Clinician Checklists.** Because the PIC was designed to allow systematic collection of parent observations of child intellectual, behavioural, and emotional status, there was some correspondence between the item content of the PIC and the teacher and clinician rating scales. Items of the teacher and clinician checklists, however, reflect child adjustment in more specific areas than are tapped by PIC items. For example, the ACH scale of the PIC has some items reflective of reading skill (e.g., reading has been a problem for my child; reading is my child's favourite pastime). The Academic Delay factor of the teacher checklist also contains items relating to reading, but these items are much more specific (e.g., below average phonics skills; below average reading comprehension), and the Academic Delay factor also contains items relating to performance in other subject areas.
Procedure
The teacher checklist was mailed to all classroom teachers and these were returned prior to the first clinic appointment. All children and their mothers were interviewed (together and separately) by clinicians (usually psychiatric residents or psychology interns supervised by senior staff). At the conclusion of the interview the mothers completed the PIC and the clinicians completed their symptom checklist and assigned a DSM-III Axis I diagnosis. If no Axis I diagnosis was assigned, the Axis II diagnosis was used as the primary diagnosis. Since all measures used in this investigation were part of the routine intake procedure at the clinic where these data were collected, subjects and their mothers were not specifically recruited for participation in this study.

In order to ensure integrity of informant response set, cases were eliminated from the analyses if their PIC profiles had elevations on the inventory validity scales, $L (T > 59)$, or $DEF (T > 69$; Lachar & Gdowski, 1979; Wirt et al. 1984). A total of 110 of 839 (13%) cases were eliminated for this reason, and all analyses were conducted with 729 cases.

Statistical Analyses
Two analyses were performed. In order to evaluate the convergence of diagnoses with the PIC, proportions of cases in each diagnostic sample that had elevated scores on PIC clinical scales were determined. Clinical- versus normal-range $T$-score cut-offs for each PIC scale (applied to both sexes 3–16 years old) were established by Lachar & Gdowski (1979), and elevated scores for ACH, DVL, FAM, and HPR are indicated by $T > 59$; the clinical-range for IS, D, SOM, WDL, ANX, and SSK is $T > 69$; and for DLQ and PSY the clinical-range is $T > 79$. Two (clinical- vs. normal-range on PIC scale) $\times$ nine (diagnostic groups) contingency table Chi-Square analyses were performed for each PIC scale to determine whether the proportion of cases obtaining clinical-range scores differed significantly across the diagnostic groups. Proportions of cases with elevated PIC scores for each diagnostic group which differed significantly from whole sample base rates were also identified using goodness-of-fit Chi-Square tests.

The second analysis was concerned with the concurrent validity of DSM-III diagnoses and the PIC. The strengths of the relation of diagnoses and PIC clinical scales to the teacher and clinician ratings were estimated with two different multivariate procedures. (Cases of the Mental Retardation and Infantile Autism groups were omitted from these analyses due to missing data.) The overall percentages of teacher and clinician checklist variance explained by DSM-III diagnoses or the PIC clinical scales were estimated using the Index of Redundancy (Stewart & Love, 1968). The Index of Redundancy is derived using canonical correlation analysis and indicates the total proportion of variance of a set of dependent variables explained by a set of predictor variables. These analyses allowed comparison of the relative ability of diagnoses and the PIC to predict these external criteria. Diagnostic group membership was represented in these analyses with dummy codes. Following these DSM and PIC analyses, the PIC scales and diagnoses were used as joint predictors of the teacher and clinician criteria. These analyses determined whether the combination of DSM and PIC information could improve prediction over that attained using either set alone.

A series of multiple correlation analyses were performed to evaluate the magnitude of relation of diagnoses and the PIC to each teacher and clinician rating dimension. The PIC clinical scales and DSM diagnoses were entered as predictor variables in separate regression equations to predict each of the 14 teacher and clinician rating scales. Also, PIC scale scores and DSM diagnosis were entered as joint predictors of each teacher and clinician dimension. Due to the large number of statistical tests performed for these analyses (14 dimensions $\times 3 = 52$), the .01 significance level was used.

RESULTS
Convergence of DSM-III Diagnoses and the PIC
Presented in Table 2 are proportions of each diagnostic group and the whole sample...
that had clinical-range scores on the PIC clinical scales. The results of the $2 \times 9$ contingency table Chi Square analyses for each PIC scale and goodness-of-fit analyses for each diagnostic group-by-PIC scales are also reported in this table.

The results in Table 2 clearly indicate that clinically elevated PIC scores were not randomly distributed across the diagnostic groups, and the results also suggest two general patterns of diagnosis-PIC scale convergence. First, relatively high proportions of several diagnostic samples obtained elevated scores on PIC scales which measure problems consistent with their diagnostic labels. Sensitivity values for PIC scales which measure problems directly indicated by DSM-III diagnostic criteria of certain diagnostic samples ranged from 43% to 93% (average = 73%). These sensitivity values (from lowest to highest) and the diagnostic samples from which they were derived are: 43% for ANX (from the Overanxious Disorder sample); 53% for D (Depression sample); 64% for DLQ (Conduct Disorder sample); 65% for SSK and 79% for PSY (Infantile Autism and Schizophrenic Disorder samples); 83% for HPR (Attention Deficit Disorder sample); 89% for ACH (Mental Retardation and Developmental Disorder samples); and 91% for IS and 93% for DVL (Mental Retardation sample). In general, the correspondence of PIC scale scores to diagnostic group membership was lowest for children assigned diagnoses reflective of internalization and was higher for children given diagnoses indicating externalization or cognitive/social impairment.

Rates of PIC scale elevations for some diagnostic groups were significantly below sample base rates, and these were consistent with expected presenting problems for these samples. For example, only 11% of the Attention Deficit Disorder cases obtained high scores on WDL, and only 7% of the Infantile Autism cases received high DLQ scores. Children who received an Adjustment Reaction diagnosis were significantly less likely to receive elevated scores on several PIC scales, including ACH, IS, DVL, and PSY.

A second pattern of diagnosis-PIC scale correspondence suggested by the results in Table 2 is that relatively high proportions of several diagnostic samples had elevated scores on PIC scales which measure behaviour problems not directly identified by their diagnostic labels. For example, a majority (53%) of the Depression cases had elevated scores on DLQ, while 50% had clinical-range scores on HPR. Similar proportions of elevated scores on these same PIC scales were received by the Overanxious Disorder cases. About half of the Conduct Disorder and Attention Deficit Disorder cases had elevated scores on the D scale.

Concurrent validity of DSM-III Diagnoses and the PIC
The results of canonical and multiple correlation analyses of the relation of diagnoses and the PIC to the teacher and clinician rating criteria are presented in Table 3. The overall multivariate percentage of variance of the teacher or clinician behaviour ratings explained by DSM diagnoses was less than 2% (0.85% of the teacher
ratings, 1.46% of the clinician ratings). The range of multiple correlation values for the individual teacher and clinician rating dimensions was from .07 (for teacher ratings of Poor Classroom Adjustment) to .20 (for clinician ratings of Hostility/Dyscontrol; average = .12). Only two of these coefficients attained significance at the .01 level. In contrast, the multivariate percentages of explained variance by the PIC scales were much higher: 15.37% of the teacher ratings and 21.32% of the clinician ratings were accounted for by the PIC scales. All 14 multiple correlations for the individual rating dimensions were significant, and ranged from .22 (for teacher ratings of Poor Self-Concept/Depression) to .57 (for clinician ratings of Antisocial Behaviour; average = .42).

The overall multivariate explained proportions of variance and multiple correlations derived from joint prediction with diagnoses and PIC scales are also presented in Table 3. Without exception these values are only marginally greater than those generated using only the PIC scales. The increments in overall explained variance using diagnosis and PIC information over using the PIC alone were only 1.58% for teacher ratings and 1.11% for clinician ratings; the range of increments for the multiple correlations was from 0 to .04. These results indicated that the unique contribution of DSM diagnoses to prediction of teacher and clinician behaviour ratings over the predictive capacity of the PIC was small.
DISCUSSION

The results of this study suggest a positive though imperfect relation between DSM-III diagnoses and the parent-informant PIC. The majority of children assigned several different diagnoses also had elevated scores on related PIC scales. The average sensitivity of several PIC scales was 73%, which is consistent with values estimated by investigators who have studied other parent-informant measures. The extent of diagnosis-PIC convergence, however, varied as a function of type and severity of child problems. That is, the proportions of children who received DSM-III diagnoses which reflect internalization (Depression or Over-anxious Disorder) and also had elevated scores on related PIC scales (D or ANX) did not differ significantly from sample base rates. On the other hand, proportions of children who were assigned diagnoses designating externalizing behaviour (Conduct Disorder or Attention Deficit Disorder) or cognitive/social dysfunction (Mental Retardation, Developmental Disorder, Infantile Autism, or Schizophrenic Disorder) and who also received elevated scores on related PIC scales were typically higher than proportions for internalizing cases.

The finding that convergence of objective parent ratings and diagnostic labels varied as a function of severity of child adjustment difficulties is consistent with results reported by Bird et al. (1987). These investigators found higher correspondence of DSM-III diagnoses and CBCL scores for children who received diagnoses which designate more severe (e.g., Infantile Autism) rather than less debilitating psychopathology (e.g., Separation Anxiety).

Results from the present and previous studies concerning the extent to which type of presenting child problem is related to clinician-parent agreement have not been as consistent. For example, Achenbach, McConaughy, and Howell (1987) conducted a meta-analysis of the results of 119 studies to evaluate the correspondence of reports from different informants (parents, clinicians, observers, teachers, peers, and children themselves) who completed rating scales about children's adjustment difficulties. While these investigators did not specifically study the agreement of diagnoses and checklist scores, their findings are nevertheless informative. Achenbach et al. reported that agreement across all informants was higher for children with emotional rather than conduct problems, the opposite pattern found in the present study. It appears at this time that severity rather than type of child problems may have the clearest relation to diagnosis-checklist convergence, and the influence of type of child problem needs to be explored.

A second pattern of results in the present study was that many cases had clinically elevated scores on PIC scales which measured problems not identified by their diagnostic labels. Considering the relatively higher concurrent validity of the PIC to DSM diagnoses in the prediction of teacher and clinician ratings, these findings suggest the PIC may have accurately identified problems missed by the diagnostic
DSM AND PIC 261

labels. At minimum these findings highlight the limitations of assigning a single diagnostic label to describe children who may have multiple adjustment problems.

Before implications of the findings of this study are discussed, review of some important caveats is in order. Some limitations concern the teacher and clinician ratings used as external behavioural criteria for this study. The PIC and the teacher and clinician scales are checklists, and the PIC may be expected to have a stronger relation to these measures than diagnoses because it shares a common measurement method. There was a characteristic of this study, however, which may have enhanced the predictive power of diagnoses relative to the PIC as regards the clinician ratings. Because clinicians assigned diagnoses after completing their own checklist, it was expected that a close correspondence would occur between diagnosis and checklist responses. This expected descriptive superiority of diagnoses may then reflect a source of bias or contamination in favour of DSM-III classification.

Another limitation of the teacher and clinician behavioural criteria is that they represent only one type of information that may be collected in the evaluation of children. It is unknown whether the predictive superiority of the PIC over diagnoses found with the teacher and clinician ratings used in this study would be replicated using other types of clinical information (e.g., structured interviews of parents). A final limitation of the checklist criteria used here is that their items refer to only a single episode of maladjustment which precipitated a referral for mental health services. A more adequate evaluation of the comparative utility of diagnostic labels and the PIC would involve examination of data relating to treatment response and outcome.

Other limitations to the generalizability of these findings concern the fact that diagnoses were assigned by residents and interns, and these trainees were given no special instruction in use of DSM-III outside of their standard clinical instruction. Thus, use of the DSM-III typology in our study was not done at an "optimal" level. Our results, however, may reflect more "typical" patterns of diagnosis-checklist convergence and relative concurrent validity in many clinical settings. Multiple Axis I diagnoses were not assigned, and this may have limited the predictive power of DSM-III diagnoses. Another limitation is that Axis III (Physical Disorders), IV (Severity of Psychosocial Stressors), and V (Highest Level of Adaptive Functioning in Last Year) were not available for our cases, and this information may have improved the predictive power of the Axis I diagnoses.

The results of this study have several implications for clinical practice with children. The relatively low concurrent validity of DSM-III diagnoses suggests that parent-informant checklists could provide information to supplement diagnostic labels. More specifically, use of such scales may be beneficial for children who receive certain types of diagnostic labels. Some DSM-III diagnostic categories (such as Pervasive Developmental Disorder or Attention Deficit Disorder with Hyperactivity) seem to designate "real" disorders. That is, they have high
interrater agreement, have been corroborated with findings from empirical studies, and offer fairly clear prognostic implications (Quay, 1986). Other DSM-III categories (such as Adjustment Reaction or Oppositional Disorder) have not been supported by results of empirical investigations and have low interrater reliabilities, raising questions about their validity (Quay, 1986). For children who are assigned such diagnoses, the availability of other sources of adjustment data (such as parent checklists) may be especially useful.

An example of the use of parent-informant rating scales to improve prognostic predictions for children who receive certain diagnoses is outlined below. Voelker, Lachar, and Gdowski (1983) administered the PIC to mothers of children assigned a diagnosis of Attention Deficit Disorder to discriminate cases showing a positive response to methylphenidate from those showing no improvement. Poor medication responders were more likely to have elevations on PIC internalizing scales (SOM, D, FAM, WDL, & ANX), while good responders obtained high externalizing scale (DLQ & HPR) scores. These findings suggest children whose behaviour meets the diagnostic criteria for Attention Deficit Disorder, and who also have high scores on PIC internalizing scales, may be reacting to environmental events (e.g., family conflict). Relying solely upon medication for treatment for such children would be inappropriate.

RESUME

Les diagnostics psychiatriques ainsi que les échelles de cotation du comportement objectif représentent des stratégies alternatives servant à décrire les troubles psychologiques des enfants. Les résultats de plusieurs études concernant la convergence des diagnostics et les échelles de cotation sont équivoques, soulevant ainsi des questions concernant la nature de la relation de ces deux systèmes d’évaluation. Le but de cette étude était d’évaluer les relations entre les diagnostics avec le DSM-III et le "Personality Inventory for Children" (PIC) auprès d’un échantillon d’enfants et adolescents démontrant des troubles comportementaux. En plus, on évaluait la capacité relative de prédiction des diagnostics du DSM-III et des pointages du PIC relativement aux cotations comportementales de professeurs et de cliniciens. Une grande proportion d’enfants auxquels on avait assigné des diagnostics précis du DSM-III obtinrent également des pointages élevés sur les échelles correspondantes du PIC, même si le recoupement n’était pas parfait. On trouva également que la covariation avec les évaluations du comportement par les professeurs et les cliniciens était plus élevée avec le PIC qu’avec les diagnostics du DSM-III. Les implications de ces résultats pour la pratique clinique avec les enfants sont discutées.

REFERENCES


First received 25 March 1987
Final revision 24 September 1987