Parent-Teacher Reliability in Rating Children on the 10-items Conners Rating Scale

T. K. Daradkeh.

The purpose of this study is to estimate the level of agreement between parental and teacher ratings of school children behaviour on the items of the short version of the Conners Rating Scale. The interrater reliability was found to be low on all items. This study casts doubts on the validity of the prevalence estimates of the hyperkinetic disorder when using the short version of the Conners scale and relying only on parents or teachers as single informants.

(Egypt.J.Psychiat.,1993,16: 52-56).

Introduction

Although hyperactivity in children is frequently diagnosed, the condition is poorly defined, and prevalence data, in this part of the world, are not available. There are many good reasons for demographic studies of hyperactivity syndrome. First, hyperactivity is the most frequent cause of referral of children to clinics, special services in schools and other professional resources. Second, there is increasing awareness that hyperactivity in children is a risk factor for problems in adolescence and adulthood (Gittelman et al., 1985; Weiss & Hechtman, 1986); hyperactivity has been implicated in delinquency, alcoholism and other social behaviour problems. Third, cross-cultural studies offer an opportunity to compare rates of occurrence, and if differences are found, this may provide information regarding the causes and pathogenesis of the disorder (Luk et al., 1988). Prevalence estimates for the hyperactivity syndrome in children vary considerably, ranging from a low of slightly over 1% to 15% (Shrag & Divoky, 1975). The most frequently quoted estimates are in the range of 5-10% (Wender, 1971), but in 40% of children seen by psychiatrists in the United States (Greenberg & Lipman, 1971), but in only 1.6% of psychiatrically diagnosed children in the Isle of Wight studies (Rutter, Tizard & Whitmore, 1970). Such large differences are very difficult to account for, but may be due to nothing more than the condition in different countries being labelled differently. Studying the prevalence of the hyperactivity syndrome in a large population demands a simple, reliable, and valid method of identifying these children. Several teachers rating scales have been developed, but one of the most widely used scales is the Conners Teacher Rating Scale (Conners, 1969). This is a 39-item behaviour symptom checklist that has five orthogonal factors, four of which are commonly used: conduct problem, inattentive-passive, tension-anxiety, and hyperactivity. A short version of the scale has also been developed and the behaviour of the child is rated on 10 items instead of 39 (Conners, 1972). Each of the items on the scale is rated using four categories: "not at all," just a little," pretty much," and "very much." Rates vary depending

Tewfik K. Daradkeh. M.D; MRC Psych, DPM. Associate Professor of Psychiatry ,Faculty of Medicine & Health Sciences ,UAE University, Al Ain. P.O. Box 17666 United Arab Emirates

on the number of sources of information used to reach a diagnosis. If parents, teachers and physicians together are requied to make a diagnosis, the prevalence figure will be lower (Lambert et al., 1978). In contrast, if a single source of information (e.g., parent or teacher) is required for a diagnosis of a single symptom, the prevalence figure tends to be higher (Lapousse & Monks, 1958). The purpose of this study is to estimate the level of agreement shown parents and teachers when rating' children's behaviour as a preliminary step to launching a large population survey.

Subjects and Method

The study was conducted in Al Ain, United Arab Emirates in 1991. The short version of the scale was chosen as a screening instrument. The study was approved by both the research and ethical committees of the Faculty of Medicine and Health Sciences and the Department of Education in Al Ain. The scale was initially translated into Arabic by the author but the final translated version was adopted after receiving comments on the translation from colleagues, bilingual English teachers and an educational psychologist who had been involved in testing and counselling of school children in Al Ain for nearly two decades. It was decided to choose one school for the study; children of kindergarten and the first four grades were rated and whose parents could read and write. To obtain the cooperation of school teachers, the school was identified by an educational psychologist familiar with that school. Two weeks before the end of the educational year 1991, 161 copies in duplicate of the scale were distributed with letters of instructions about how to rate a child's behaviour on the 10-item Conners Rating Scale (161 parents gave verbal consent to school authorities to participation). Teachers and parents were requested to rate the child's behaviour independently. Each item was rated on a 4-point scale (0=not at all; 1= just a little; 2=pretty much;3=very much). Tests of associations were used to assess the significance of correlation on each item of the scale; Kappa and Repeatability Reproducibility Measurements were also applied to estimate the level of agreement. The latter procedure was applied on the total scores since the distribution of total scores approximate that of normal distributions. Wilcoxon Matched Pair Test was applied to assess the significance of differences between parental and teachers' ratings of each item; Paired t-test was used to assess the differences between the means of total scores. We also tested the hypothesis that high total rating (one standard deviation above the mean) of a child by one rater, would be predictive of their ratings by the other rater, using the interactive outlier regression analysis.

Results

Ninety per cent of the scales were returned (145 copies in duplicate), but only 128 pairs were filled completely (85 boys and 43 girls). Statistical analysis was carried out on these 128 children. This represents 80% of children whose parents could read and write. Table 1. shows the distribution of children by items of behaviour and by raters and reveals that individual symptoms are not rare among the sample. It reveals low correlational, agreement levels between parental and teachers' ratings on all the items of the scale. It also shows that parents, when compared with teachers, overrated children on the item of restlessness while teachers overrated children on the items of inattention, fails to finish things and mood changes quickly. There were no significant differences between the means of the total parental and teachers' ratings (t=0.19, N.S.). Product moment correlation was significant but low

Table 1
Distribution of 128 Children by Rating Categories and by Raters

	Scoring Categories									
	P 0 %	T	P 1 %	T	P 2 %	T	P 3 %	T	K	kendall
1 Restlessness*	2	4	13	33	52	39	32	24	0.21	0.27
2 Excitabe	24	26	43	44	20	23	13	7	0.28	0.33
3 Disturbs other	45	33	25	38	25	24	5	5	0.20	0.34
4 Fails to finish things**	64	55	26	42	7	9	3	4	0.26	0.35
5 Fidgety	23	18	30	35	29	38	18	9	0.26	0.35
6 Inattention**	53	35	31	43	10	18	6	4	0.29	0.43
7 Demands must be met	31	20	42	56	13	18	14	6	0.26	0.27
8 Cries often	44	49	33	32	11	15	12	4	0.16	0.15
9 Moody**	52	37	36	45	7	16	5	2	0.26	0.28
10 Temper outbursts	62	56	27	35	5	7	6	2	0.25	0.28

*P = Parents; T = Teachers K = Kappa Kendall = tau B
* = Overrated by parents
** = Overrated by teachers

between the total scores (r-0.33, P<0.01). Repeatability Reproducibility Measurement of the total scores reveals that 70% of the variability was due to children and 30% of the total variability were due to raters. The latter component should be less than 10% to consider the ratings as reproducible. There were 16 children whose total parental ratings were one standard deviation above the mean but their total ratings did not predict their teachers total ratings (r=-0.46, N.S.). However, teachers' ratings revealed that there were 15 children whose total scores were one standard deviation above the mean; their ratings significantly predicted the parental ratings of the same children (r=0.64, p<0.01). The interactive outlier regression revealed that there was one outlier and by excluding it from the analysis, the product moment correlation coefficient rose to 0.78. Examining the total parental and teaches total scores Stem and Leaf procedurre revealed that there were 5 extreme parental total scores and 2 extreme teachers total scores but only one child was identified as being extremely rated by both parents and teachers.

Discussion

To this study, the sample size was adequate, since correlational analyses on larger samples may reveal significant correlation when the magnitudes of such correlation are terribly low. We postulate that the low level of agreement on all measures between parental and teachers' ratings of schoolchildren may be due to a combination of factors. Children behave differently in different situations; observers differ in their degree of tolerance of "abnormal" behaviour and some misunderstand items. The findings that a high score assigned by teachers significantly predicted parental ratings, but the reverse was not true, suggests that teachers may be more accurate in their judgments in ratings' schoolchildren. It has been reported that parental ratings show no sig-

nificant relationships to either objective measures or ratings done outside the home (Langhorne et al., 1975; Rapoport & Benoit, 1976; Routher & Schroeder, 1976). Riddle and Rapoport (1976) examined the stability of teachers' and parent ratings over time. They found that teachers' ratings over a 2-year period were more stable than parent ratings even with different teachers doing the ratings, and even when the child had changed schools. When interrater reliability and test-retest reliability of teachers' ratings of schoolchilden on the Conners Teachers Rating Scale were examined using thousands of children, the results were encouraging (Trites et al., 1983; Quay et al., 1979; Luk et al., 1988). Due to situational influences on children's behaviour, and different conceptions among observers concerning what is "abnormal," the parent-teacher reliability in ratings' schoolchildren's behaviour is expected to be low. It is noteworthy that when both parents and teachers are required to assign extreme scores the figure was five times lower than the figure obtained by parental ratings alone. These findings cast doubts on the validity of data obtained concerning the prevalence of the disorder when relying especially on parents as single source of obtaining information about the extent of the disorder. It is also of interest to mention at this point that if we assume that the symptoms should be pervasive in all situations the proportion of children with extreme scores is much lower than the proportions of children judged as being extremes by either parents or teachers. We have estimated the scale reliability as used by parents and teachers and we found that the Cronbach value was high (0.77 and 0.73 respectively). This suggests that over 70% of the observed scores were accurate as viewed by parents and teachers or the items discriminated reliably between children; though there was a low agreement between parental and teacher's ratings. The consistency in parental and teachers' ratings though poorly correlated implies that if both ratings are considered in labelling a child as hyperactive, the validity of the prevalence estimates will be higher but the rate will be lower. If pervasiveness of the symptoms irrespective of the situations is adopted as criteria for the diagnosis, it is highly likely that the obtained figure of prevalence rate will be close to the real rate. Despite the findings that teaches may be more reliable in rating children behaviour; relying on parents or teachers alone as a single source of information in estimating the extent of the hyperactivity syndrome carries the risk of obtaining unreal prevalence estimates. It seems also that the scale in its present format is unlikely to have a reasonable validity.

References

Conners, C.K. (1969). A teacher rating scale for use in drug studies with children. *Amer. J. Psychiat.*, **126**: 884-888.

Conners, C.K. (1972). Symposium: Behaviour modification by drugs 11. Psychological effects of stimulant drugs in children with minimal brain dysfunction. *Paediat*, **49**: 702-708.

Greenberg, L., Lipman, R. (1971). Pharmacotherapy of hyperactive children. Current Practices. *Clin. Proc. Child Hosp*, **27**: 101-105.

Gittelman, R. Mannuza, S. Shenker, R. et.al. (1985). Hyperactive boys almost grown-up. *Arch. Gen. Psychiat.*, **42**: 937-947.

Lambert, N.M. Sandoval, J. Sassone, D. 1978). Prevalence of hyperactivity in elementary school children as a function of social system definers. *Amer. J. Orthopsychiat*, **48**: 446-463.

Lapouse, R. Monks, M. (1958). An epidemiological study of behaviour characteristics in children. *Amer. J. Public Health*, **48**: 1134-1144.

Luk, S.L., Leung, P.W.L., Lee, P.L.M. (1988). Conners Teacher Rating Scale in Chinese children in Hong Kong. *J. Child Psychol. Psychiat.*, **29** (2): 165-174.

Quay, H.C. (1979). Classification. In H.C. Quay and J.S. Werry (eds) Psychopathological Disorders in Childhood (2nd ed). New York: Wiley.

Rapoport, J.L., Benoit, M. (1975). The relation of direct home observation to the clinic evaluation of hyperactive school age. *J. Child Psychol. Psychiat.*, **16**: 141-147.

Routher, D.K., Schroeder, C.S. (1976). Standardized playroom measures as indices of hyperact. *J. Abnor Child Psycho.* **4**: 199-207.

Riddle, K.D., Rapoport, J.L. (1976). A 2-year follow-up of 72 hyperactive boys, Classroom behaviour peer acceptance. *J. Nerv. Ment. Dis.*, **162**: 126-134.

Rutter, M., Tizard, J., Whitmore, K. (1970). Education, Health and Behaviour Psychological and Medical Study of Childhood Development. New York: Wiley-Interscience.

Schrag, P., Divoky, D. (1975). The myth of the hyperactive child. New York: Pantheon Books.

Trites, R.L., Laprade, K. (1983). Evidence of an independent syndrome of hyperactivity. *J. Child Psychol. Psychiat*, **24**: 573-586.

Weiss, G., Hechtman, L. (1986). Hyperactive children grown up: Empirical Findings and Theoretical Considerations. New York: Guilford Press.

Acknowledgement: I thank all those who contributed in the translation of the scale and to parents and teachers who cooperated in the study.

La Fiabilité de l'Evaluation des Enfants par le "10 - points Conners série d'évaluation

Le but de cette étude est d'éstimer le niveau de l'accord entre les évaluations des parents et ceux des professeurs concernant le comportement des étudiants de l'âge scolaire sur les points de la version courte de la série d'évaluation de Conners. La fiabilité des inter-évaluteurs a été trouvée bas en ce qui concerne tous les points.Les moules de cette étude doutent la validité de l'évaluation des troubles hypercinétique en utilisant La version courte de la serie d'évaluation de Conners, et en comptant seulement sur les parents et les professeurs comme informants.

معدل ثبات استبيان كونورز المختصر بين المدرس والأبوين لدى تقييم الأطفال

تهدف هذه الدراسة إلى تقدير مدى الاتفاق بين الأبوين والمدرس فى تقييم سلوك أطفال المدارس على مفردات النسخة المختصرة لاستبيان كونورز. وقد تبين أن معامل الثبات بين المقيمين منخفض على كل المفردات فى الاستبيان، وبالتالى تثير هذه الدراسة الشكوك حول مصداقية معدلات انتشار اضطراب زيادة الحركة المعروفة والمستفادة باستخدام النسخة المختصرة لمقياس كونورز، كما تشكك فى جدوى الاعتماد على تقييم منفرد للطفل بواسطة المدرس أو الوالدين.