Seizure 2003; **12:** 261–267
doi:10.1016/S1059–1311(02)00275-3

The Washington Psychosocial Seizure Inventory (WPSI): psychometric evaluation and future applications

CHIH-HUNG CHANG [†,‡] & SARAH GEHLERT [§]

[†]*Center on Outcomes, Research and Education (CORE), Evanston Northwestern Healthcare, 1001 University Place, Evanston, IL 60201, USA;* [‡]*Institute for Health Services Research and Policy Studies, Northwestern University, Evanston, IL 60201, USA;* [§]*School of Social Service Administration, The University of Chicago, Chicago, IL, USA*

Correspondence to: Dr Chih-Hung Chang, Ph.D., Center on Outcomes, Research and Education (CORE), Evanston Northwestern Healthcare, 1001 University Place, Suite 100, Evanston, IL 60201, USA.
E-mail: chchang@northwestern.edu

The Washington Psychosocial Seizure Inventory (WPSI) clinical scales were developed via an empirical item selection approach and have been used widely to measure aspects of psychosocial functioning of patients with epilepsy. However, these empirically derived clinical scales have not been assessed psychometrically using a modern item response theory-based model. The goals of this study were to: (1) evaluate how items in each clinical scale performed in such a way as to represent the underlying constructs being measured; and (2) derive a shorter version while maintaining measurement precision. WPSI item response data from 145 adults with epilepsy collected for an evaluation study of an intervention to pact negative attributional style in epilepsy were used. The dichotomous Rasch model suitable for the true–false response choices was used to analyse each clinical scale separately. Most items within each scale fit the measurement model well, with very few exceptions. All items, therefore, were retained. A method, based on computerised adaptive testing (CAT), is offered for shortening the WPSI using a psychosocial outcomes item bank derived from the study. Individuals' specific levels of functioning are used to derive measures of their psychosocial functioning with a minimum number of items.

Key words: Washington Psychosocial Seizure Inventory; item response theory; psychometrics; computerised adaptive testing.

INTRODUCTION

The Washington Psychosocial Seizure Inventory (WPSI)[1] was for many years the only standardised psychological instrument designed specifically for testing adults with epilepsy. As such, it has been used by many professionals in the field of epilepsy to identify psychosocial problems associated with the condition. To date, over 60 articles have been written on the WPSI (see Dodrill[2] for a review to that date), and it has been translated into several languages.

The WPSI was developed in 1978 by researchers at the Epilepsy Center of the University of Washington, School of Medicine. Dodrill and Batzell[3] states that after considerable work in the area, including pilot studies, he and his colleagues were able to identify the seven psychosocial areas which later became the scales of the WPSI (an overall index was also included, thus yielding eight scales). Next, an item pool was created for each scale, "with items tested for clarity and for minimal overlap" (Dodrill *et al.*[1], p. 124).

The authors state that the final scales of the WPSI were determined on an empirical basis[1]. They say that this approach was chosen, "because it has proven very successful with patient groups with such tests as the MMPI" (Dodrill *et al.*[1], p. 127). The empirical approach utilised differs from that used by the authors of the MMPI, however. In the latter, items were selected for the test that were able to distinguish known groups, that is to say they reflected statistically significant differences between a criterion group of subjects with known pathology and control subjects[4]. On the WPSI, items were chosen which correlated significantly with ratings obtained from professionals on what was called the Professional Rating Sheet. Thus, Professional Rating Sheet scores were used as criteria

or external validators. Dodrill *et al.*[1] conclude that, "in this fashion, each scale was composed on a purely statistical basis regardless of the content of the items" (p. 129). Three validity scales modelled after those of the MMPI[5] were added to the original scales to identify unreliable responses.

The authors of the WPSI[1] report test–retest reliability coefficients ranging from 0.66 to 0.87. Internal consistency was assessed by the split-half method with coefficients of from 0.68 to 0.95 obtained[1]. Criterion validity was established by comparing scores on the WPSI with ratings by professionals, yielding coefficients from 0.50 to 0.75[1].

Langfitt[6] evaluated the psychometric properties of the WPSI with a sample of 71 patients and compared the group's scores on the WPSI to those on two measures of quality of life, the Epilepsy Surgery Inventory-55 (ESI-55)[7] and the Sickness Impact Profile (SIP)[8]. Cronbach's alpha coefficients ranged from 0.62 to 0.94 for the WPSI clinical scales. Face validity was established by comparing how two raters classified the WPSI's items as belonging to the 10 domains of quality of life outlined by Hermann[9]. A Kappa coefficient[10] of 0.58 was obtained, indicating low to moderate agreement between raters. A method of measuring content validity was used in which, using the two raters' results, the 10 domains were collapsed into three (physical functioning, psychological functioning, and role activities) and the WPSI's scores on the three domains were compared to similar scores for the ESI-55[7] and the SIP[8] in a 3×3 table using χ^2. A χ^2 of 60.1 ($P < 0.0001$) was obtained, suggesting that the three measures differed markedly in the distribution of their items across domains. Construct validity was measured using the multitrait–multimethod technique[11], with the three instruments as methods and domains scores on emotional, interpersonal, and vocational adjustment as traits. Coefficients of 0.47, 0.50, and 0.77 were calculated for the WPSI and ESI-55 vocational adjustment, interpersonal adjustment, and emotional adjustment, respectively. Coefficients of 0.55, 0.61, and 0.62 were obtained for the same domains of the WPSI and SIP. Criterion validity was assessed using general linear model, one-way ANOVA. Total scores for the measures were compared to known group differences via relative efficiency ratios[12] using the overall psychological functioning scale on the WPSI as a reference measure. Relative efficiency ratios of 2.63 and 2.16 for the ESI-55 and SIP, respectively, were determined. Langfitt[6] concluded that the EPI-55 and SIP were preferable to the WPSI for measuring quality of life.

That the WPSI did not fare well psychometrically when criterion validity was assessed using quality of life measures is not necessarily surprising and can be supported using conceptual arguments. Although, the WPSI's use was extended to measuring quality of life during its second decade[2], the move was met with some controversy. Selai and Trimble[13] make the argument that an instrument designed to measure quality of life should have patient input on item selection, because "patients can best identify the issues relevant to their quality of life" (p. 335). That patients do not appear to have been involved directly in item selection runs counter to the current philosophy of quality of life that patients can best identify issues relevant to their own lives[14, 15]. Selai and Trimble[13] note additional shortfalls of the WPSI; that it is long and that its administrators must be trained. Hauser[16] argues against the WPSI's use with a wide range of people with epilepsy, saying that it is best used for what it was originally designed, namely screening patients at referral centres for severity of psychosocial problems.

The present study has two aims. The first is to provide an objective assessment of the WPSI's psychometric properties using methods that go beyond traditional means of testing validity and reliability. Although the validity and reliability of the instrument have been addressed in the past, largely by its authors, the conceptualizations of validity and reliability used were limited. Extensive assessment performed by outside, and therefore presumably more objective, sources would allow more confidence in the WPSI's use. We applied the dichotomous Rasch model[17–21], one-parameter logistic model, to the clinical scales of the WPSI to gain a better understanding of what its scales are measuring.

The study's second aim is to explore solutions to two criticisms of the WPSI outlined earlier; that it is long and that it must be interviewer administered. The Rasch model will be used to determine whether a shorter form of the WPSI can be developed that is at least psychometrically sound as the current version. Validity of each clinical scale will be determined by examining how the items perform to represent each psychosocial functioning construct. Items found not to be contributing significantly to the WPSI's constructs will be retooled or removed and the instrument re-calibrated. A measure of reading difficulty termed a Lexile (developed for MetaMetrics of Durham, NC)[22] will be calculated for each WPSI item to measure its reading difficulty. This will be done to determine the requisite reading skills needed to comprehend the WPSI items.

METHOD

Participants

The data, from 145 adults with epilepsy living in the City of Chicago, were collected for an evaluation

study of an intervention to impact negative attributional style in epilepsy[23]. Participants were recruited via letters of solicitation passed out to patients checking into hospital-based comprehensive epilepsy clinics over a 2-month period. Potential participants were given a telephone number to call via which they were screened to see if they met the study's inclusion criteria, age 18 years or older, no history of mental retardation, and a diagnosis of epilepsy made by a physician.

Measure and procedures

The WPSI contains 132-item true–false questions. Test items are all short and written in simple language and as a rule take only 15–20 minutes to complete. Its eight scales capture the types of psychosocial concerns commonly seen in epilepsy: family background (11 items), emotional adjustment (34 items), interpersonal adjustment (22 items), vocational adjustment (13 items), financial status (7 items), adjustment to seizures (15 items), medicine and medical management (8 items), and overall psychosocial functioning (57 items). These scales are scored dichotomously.

Individual participants were given the WPSI under the supervision of Masters-level students trained in administering the instrument. Demographic and clinical information were gathered at the same time. WPSI data were collected before and after the intervention to measure overall psychosocial functioning. Pre-test data were used in the present study.

Demographic and clinical information was collected to assess the degree to which participants in the study represented what we know about adults with epilepsy in the United States. Demographic variables measured were age, gender, marital status, and employment. Clinical variables were age at onset of seizures and seizure type.

Analysis

Rasch model

The dichotomous Rasch model[17, 22], suitable for the true–false response choices of the WPSI items, was used to analyse each clinical scale separately. The WINSTEPS computer program[24] was used for Rasch analyses.

The dichotomous Rasch model was originally devised to construct ability tests based in the logistic response model that fully warrants objective measurement[17]. It specifies through log-odds that the probability of response of person n to the item i is governed by the location B of the subject (person measure) and the location of D of the item (item calibration) along a common measurement continuum:

$$\log\left[\frac{P_{ni1}}{P_{ni0}}\right] = B_n - D_i$$

where P_{ni1} is the probability of a "true" response to the item i and P_{ni0} is the probability of a "false" response. When $B_n > D_i$, there is more than a 50% chance of a "true" response. When $B_n = D_i$, the chance for a "true" response is 50%. When $B_n < D_i$, the probability is less than 50%. Each facet in the model (B, D) is a separate parameter. For instance, a person with more ability than another has a higher probability of completing a task, and thus has a higher measure.

This linear relationship between the log-odds, patient level and WPSI item locations allows us to establish an underlying dimension on a continuum, along which respondents and items can be jointly placed, ordered, and compared (see Fig. 1). It provides interval scale person measurement and item calibration, in which the person measures are independent of the items used to estimate them, and the item calibrations are independent of the distribution of person measures. It is this property that allows us to validate the difficulty ordering of each clinical scale (from less to more) and establish objective definitions of the WPSI

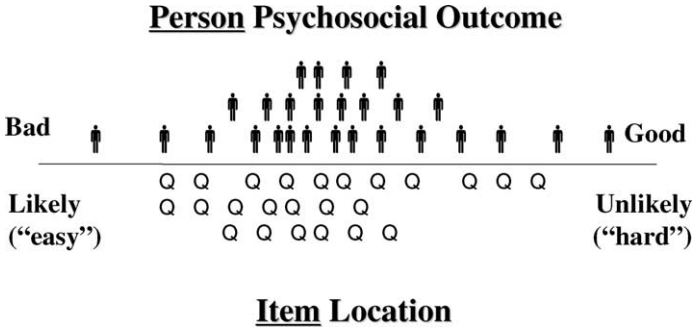

Fig. 1: Persons and items jointly placed and ordered on the same psychosocial outcome continuum based on the Rasch dichotomous model.

scales. This allows us to examine the content and degree of magnitude of questions along the continuum, and thereby better understand and define what is being measured.

Item statistics

In order to determine the spacing of each estimate, an associated standard error was calculated for each item. Other separation indices, indicating the extent to which items and persons identify a useful variable line, were also calculated. The person separation index gives the sample standard deviation in standard error units; it equals the square root of the ratio of true variance of person measures to the error variance due to person measurement imprecision. The item separation index indicates how well items spread along the variable line by giving the item standard deviation in calibration error units. Reliability (R) of person and item separation is provided by considering the relationship between R and separation (SEP) as:

$$R = \frac{\mathrm{SEP}^2}{(1 + \mathrm{SEP}^2)}$$

Person separation is the ratio of an estimate of the standard deviation (from which bias due to measurement error has been removed) to the root mean square standard error over persons. The person separation statistic provides the number of root mean square units differentiating persons and indicates how well the items can separate persons. In practice, one seeks a separation index of at least 2.0, as this would reflect the ability of a given instrument to differentiate three distinct groups of people (e.g. high, medium and low) on constructs such as aspects of psychosocial functioning in epilepsy.

Fit statistics

The Rasch model specifies that all items in a test or scale measure the same trait, which means that the test or scale is one-dimensional. Several indices are available to examine the fit of data to the model. These fit statistics provide evidence regarding the validity of the Rasch model and indices of the coherence of items to a single construct. The outfit mean square (MNSQ) statistic provides an index of the mean square residual difference between the observed and expected response patterns. The infit MNSQ statistic is a weighted mean square with weights proportional to the variance of the squared residuals and is less sensitive to outlying unexpected responses. The expected value for both of these mean square statistics is 1.0.

In the case of persons, fit statistics allow us to identify participants whose responses are somewhat unexpected or unusual. Clearly, misfitting observations can be identified for diagnostic purposes and corrective action taken when needed. In the case of items, fit statistics identify the extent to which an item contributes to the unidimensionality and additivity of the scale as a whole. If the fit is unacceptable, the data can be improved by eliminating the items that show a poor fit (and in theory, the persons that have a poor fit, as well). Thus, after eliminating misfitting items, the refined scale is assumed to be one-dimensional. It should be noted that the misfitting items identified by Rasch analysis should be examined carefully. They can either be kept for conceptual or practical reasons or set aside and analysed separately for clinical interest.

Lexile analysis

Lexiles are based on sentence length and word frequency in popular literature, with higher values indicating higher levels of reading difficulty. Since a Lexile measure refers to text readability, its primary utility is to forecast what happens when readers confront text.

The Lexile values of each WPSI item were obtained using the Lexile Analyser program. The possible range of these values is from somewhat below 0 to 1600. A Lexile value of 300 corresponds to a second-grade level of reading difficulty (UK year 2), 400 to third-grade level (UK year 3), with 1300 being equivalent to a high-school-graduate level (UK year 12).

RESULTS

Demographic and clinical characteristics of the sample

Mean age of the sample at testing was 39.6 years (SD = 10.9) (range 18–69 years). About the same percentage (50%) was male and female; 42.9% Black, 45.1% White, and 13% other. Fifty-six percent of participants said that they were single, 25.6% married, 15.2% separated or divorced, and 3.2% widowed. Percentage frequencies for employment status were: 59.8% were unemployed, 27.6% employed, 4.7% retired, 3.9% student, and 3.9% homemaker. The mean age at onset of seizures was 15.9 years (SD = 11.2). The percentage frequency of types of epilepsy in the sample was 67.7% for generalised epilepsy and 32.3% for partial epilepsy.

The frequency of categories of most demographic and clinical variables measured was consistent with national norms for persons with epilepsy. The study sample was comparable to the population of persons with epilepsy as a whole with respect to gender ratio and marital status[25, 26]. The sample approximated

Table 1: List of misfitting items and their item fit statistics by clinical scale.

Scale/item	Outfit MNSQ	Infit MNSQ
Family background		
24. Are your free from problems with your family?	1.20	1.80
Emotional adjustment		
110. Do you have a chance for vocational advancement?	1.27	1.42
59. Have you engaged in sexual practices which cause you concern or worry?	1.22	1.47
37. Do you need immediate psychiatric care?	1.10	1.36
Interpersonal adjustment		
44. Do you feel most people are phony or insincere?	1.18	1.42
79. Are you comfortable being alone despite possible seizures?	1.20	1.30
Vocational adjustment		
102. Would you like to be closer to public transportation?	1.26	1.42
15. Do you often have trouble sleeping?	1.30	1.40
Financial status		
42. Do you usually feel rested when you awake?	1.72	3.35
Adjustment to seizure		
125. Have you ever been teased because of your seizures?	1.34	1.65
92. Are you fearful of accidents?	1.29	1.58
Medicine and medical management		
63. Do you often feel restless?	1.12	1.51
62. Do you frequently have trouble remembering to take your medications?	1.32	1.39
72. Do you feel your seizures are being controlled as well as they can be?	1.20	1.31
Overall psychosocial function		
124. Are you satisfied with your employment situation?	1.21	1.50
132. Did you feel secure in the home in which you grew up?	1.15	1.39
2. Were you usually happy as a child?	1.13	1.31

MNSQ, mean square.

the general population of persons with epilepsy in terms of age at onset of seizures and seizure type experienced[26]. The study population differed from the population of persons with epilepsy as a whole in its higher rate of unemployment and its overrepresentation of Black and under representation of White participants[26].

Rasch analysis

The ideal mean square (MNSQ) value is 1.0 (i.e. observed variance = predicted variance). However, limited unexpected variance is allowed. For this study, an item with no more than 30% unexpected variance than the model predicted (i.e. MNSQ values $\leq$ 1.3) was considered evidence that the same construct was measured in that item as in the rest of items in the scale (unidimensionality). Table 1 lists items that did not fit the measurement model by clinical scale. As can be seen, a few items (Items 15, 42, 125, and 62) had both infit and outfit MNSQs greater than 1.3 cut-off criterion, indicating that these item did not work well with the rest of the items in the clinical scale to which they belonged.

To appraise the extent to which the questions on each clinical scale fully captured the range of aspects of psychosocial outcomes experienced by the

Table 2: Item statistics for the original and shortened WPSI clinical scales.

Scale	Original (shortened)				
	Number of items	Person		Item	
		Separation	Reliability	Separation	Reliability
Family background	11 (10)	3.91 (1.95)	0.94 (0.79)	1.02 (0.89)	0.51 (0.44)
Emotional adjustment	34 (31)	5.09 (5.03)	0.96 (0.96)	2.35 (2.37)	0.85 (0.85)
Interpersonal adjustment	22 (20)	3.49 (3.64)	0.92 (0.93)	1.67 (1.63)	0.74 (0.73)
Vocational adjustment	13 (11)	2.28 (2.12)	0.84 (0.82)	1.34 (1.29)	0.64 (0.63)
Financial status	7 (6)	4.87 (5.46)	0.96 (0.97)	0.84 (0.54)	0.41 (0.23)
Adjustment to seizure	15 (13)	3.11 (3.40)	0.91 (0.92)	1.62 (1.54)	0.72 (0.70)
Medicine and medical management	8 (5)	3.28 (2.98)	0.91 (0.90)	0.96 (0.08)	0.48 (0.01)
Overall psychosocial function	57 (54)	4.38 (4.39)	0.95 (0.95)	2.82 (2.84)	0.89 (0.89)

participants, we also examined the person and item separation and reliability indices. Indices from both original item compositions for each clinical scale and a shortened version (dropping those potential misfitting items) were compared. The results shown in Table 2 indicate that dropping those misfitting items identified in Table 1 failed to improve person or reliability indices significantly, suggesting that all items could be retained.

The mean Lexile measure for all 132 items was 419.55 with a SD of 215 (range = 0–1000). This indicates that a minimum of a third-grade reading level was needed to understand most of the questions.

DISCUSSION

All WPSI clinical scales were found to be acceptably uni-dimensional. Dropping potential misfitting items did not improve measurement property dramatically, suggesting that the items in each scale worked well together to measure specific underlying constructs. The reading level required to comprehend WPSI questions was found to be very minimal, supporting claims that it could be self-administered.

Although one might argue against the suitability of the WPSI for assessing quality of life, its long history of screening patients for psychosocial problems cannot be overlooked. Screening patients for psychosocial problems can improve treatment outcomes by recognising the salience of nonsomatic aspects of the condition that can stall progress, even after seemingly optimal approaches to medical management have been developed and implemented. The present study lends support to the soundness of the WPSI for measuring psychosocial functioning in the domains that it purports to measure. The instrument held up well to rigorous psychometric evaluation using a state-of-the-art measurement model; neither deletions nor additions of items seemed warranted.

Although the WPSI is sound psychometrically and can be self-administered safely to persons with at least a third-grade reading level, its critics are correct that its length detracts from its appeal as a screening tool. Because time-consuming test approaches are not feasible in busy clinic settings, having available a very brief, practical assessment of clinically meaningful, responsive data is imperative. The fixed-format computerised version of the WPSI currently in use presents questions sequentially (one at a time) as they appear in the traditional paper-and-pencil (P&P) format and questions are scored in real-time (right after questionnaire administration). This method fails to reduce patients' or administrators' burdens, however. Patients still need to answer all the 132 items before scale scores can be obtained, for example.

Computerised adaptive testing (CAT)[27], an item response theory-based methodology, coupled with advanced computer technology, offers some attractive features to reduce patients' burden and to arrive at same level of measurement precision compared to P&P administration. In CAT, items are selected from an item bank (e.g. WPSI item pool) on the basis of patients' responses to previously administered items. This process utilises an algorithm to: (1) estimate a person's level of the trait being measured (e.g. degree of interpersonal adjustment) based on persons' responses to the items; (2) select the next best targeted item; and (3) administer the test considering issues such as content coverage and test length. CAT targets the "difficulty" of items to the psychosocial functioning "level" of the responding patient. With CAT, patients need only answer a limited number of items in order to obtain a measure that accurately estimates what would have been obtained had the entire set of items been administered. For instance, if a patient answered "yes" to a question like "I can walk for a mile", subsequent questions like "I can walk for a block" would be unnecessary since they would not provide any additional, useful information. By dynamically tailoring the test to the "level" of physical functioning of individual patients, uninformative items can be eliminated.

To date, CAT has been used successfully in educational, licensing and achievement testing, personality assessment, military personnel selection, and headache assessment, and CAT can reduce test length without loss of precision. Utilising the WPSI item pool via CAT to develop a dynamically tailored and yet shortened version of the instrument obviates a major obstacle to the WPSI's usefulness, namely its length, and allows an effective tool for screening psychosocial problems to be used in epilepsy treatment centres without sacrificing clinical efficiency. Once the program has been developed and computer versions installed in the clinic, staff time can be saved, because patients can complete the screening process while they wait to be seen. Patients need only be directed to the computer screen, where simple step-by-step instructions are available, and CAT can automate administration, data recording, scoring and reporting, thus making clinical application more practical and efficient.

Clinical screening for psychosocial problems of persons with epilepsy is now considered a standard of optimal treatment. After years as the mainstay of psychosocial screening, the WPSI has lost ground to instruments that have proved to be more streamlined. We argue that the WPSI is a strong, effective instrument that merits a continued role in epilepsy testing. CAT offers a simple, modern means of bringing the instrument up to date by: (1) tailoring measurement to

individual patients; and (2) allowing testing and scoring to occur in real-time, thus allowing more time for treatment as time for diagnosis is shortened. Empirical work on the use of CAT in epilepsy care is both warranted and needed.

ACKNOWLEDGEMENT

This research was supported by grants from the Epilepsy Foundation of America and the Sarah Lind Fund of the University of Chicago.

REFERENCES

1. Dodrill, C., Batzel, L., Queisser, H. and Tempkin, N. An objective method for the assessment of psychological and social problems among epileptics. *Epilepsia* 1980; **21**: 123–135.
2. Dodrill, C. Psychological assessment in epilepsy. In: *Epilepsy: A Handbook for the Mental Health Professional* (Ed. H. Sands). New York, Brunner/Mazel, 1982: pp. 111–132.
3. Dodrill, C. and Batzell, L. The Washington Psychosocial Seizure Inventory and quality of life. In: *Epilepsy and Quality of Life* (Eds M. R. Trimble and W. E. Dodson). New York, Raven Press, 1994: pp. 109–122.
4. Cohen, R., Montague, P., Nathanson, L. and Swerdlik, M. *Psychological Testing: An Introduction to Psychological Test and Measurements*. Mountain View, CA, Mayfield, 1988.
5. Dahlstrom, W., Welsh, G. and Dahlstrom, L. *MMPI Handbook*, Vol. I. Minneapolis, MN, University of Minnesota Press, 1972.
6. Langfitt, J. T. Comparison of the psychometric characteristics of three quality of life measures in intractable epilepsy. *Quality of Life Research* 1995; **4**: 101–114.
7. Vickery, B. G., Hays, R. G., Graber, J., Rausch, R., Engel, J. Jr and Brook, R. H. A health-related quality of life instrument for patients evaluated for epilepsy surgery. *Medical Care* 1992; **30**: 299–319.
8. Bergner, M., Bobbitt, R. A., Carter, W. B. and Gibson, B. S. The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care* 1981; **19**: 787–805.
9. Hermann, B. P. Quality of life in epilepsy. *Journal of Epilepsy* 1992; **5**: 153–165.
10. Fleiss, J. L. *Statistical Methods for Weights and Proportions*. New York, Wiley, 1981.
11. Campbell, D. T. and Fiske, D. W. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin* 1959; **56**: 81–105.
12. Liang, M. H., Larson, M. G., Cullen, K. E. and Schwartz, J. E. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis and Rheumatism* 1985; **28**: 542–547.
13. Selai, C. E. and Trimble, M. R. Quality of life assessment in epilepsy: the state of the art. *Journal of Epilepsy* 1995; **8**: 332–337.
14. Cella, D. F. Quality of life: the concept. *Journal of Palliative Care* 1992; **8**: 8–13.
15. Guyatt, G. H., Feeny, D. H. and Patrick, D. L. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; **118**: 622–629.
16. Hauser, W. A. The distribution of mild and severe forms of epilepsy. In: *Epilepsy and Quality of Life* (Eds M. R. Trimble and W. E. Dodson). New York, Raven Press, 1994: pp. 249–258.
17. Rasch, G. *Probability Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmarks Paedogogiske Institute, 1960.
18. Rasch, G. On general laws and the meaning of measurement in psychology. In: *Proceedings of Fourth Berkeley Symposium on Mathematical Statistics*, Vol. 4. Berkely, CA, University of California Press, 1961: pp. 321–334.
19. Rasch, G. An item analysis which takes individual difference into account. *British Journal of Mathematical and Statistical Psychology* 1966; **19**: 49–57.
20. Rasch, G. An individualistic approach to item analysis. In: *Reading in Mathematical Sciences* (Eds P. Lazarsfeld and N. V. Henry). Chicago, Science Research Association, 1966: pp. 89–107.
21. Wright, B. D. and Stone, M. H. *Best Test Design*. Chicago, MESA Press, 1979.
22. Stenner, A. J., Horabin, I., Smith, D. R. and Smith, M. Most comprehension tests do measure reading comprehension: a response to McLean and Goldstein. *Phi Delta Kappan* 1988; **69**: 765–769.
23. Gehlert, S. Cognitive restructuring for psychosocial problems in epilepsy. *Epilepsia* 1995; **35** (Suppl. 7): S140.
24. Wright, B. D. and Linacre, J. M. *WINSTEPS: Rasch Analysis for All Two-Facet Models* (*Computer Software*). Chicago, MESA Press, 1998.
25. Lectenberg, R. *Epilepsy and the Family*. Cambridge, Harvard University Press, 1984.
26. Hauser, W. A. and Hesdorffer, D. C. *Epilepsy: Frequency, Causes and Consequences*. New York, Demos Publications, 1990.
27. Wainer, H. *Computerized Adaptive Testing: A Primer*. Hillsdale, NJ, Lawrence Erlbaum, 1990.