Qualitative Issues in Developing Educational Diagnostic Instruments and Assessment Procedures for Diabetic Patients

RICHARD A. WINDSOR, JEFFREY ROSEMAN, GEORGE GARTSEFF, AND KATHARINE A. KIRK

With one exception, psychometric analyses of the quality of instrumentation and educational assessment procedures for diabetic patients have not been reported in the literature. Following an extensive internal review process, a pilot test of 56 diabetic patients found that the newly developed instrument had a high degree of internal consistency for the major indexes, 0.89 and 0.85, respectively. An item analysis found individual questions to be of an acceptable quality. An analysis of interrater reliability of patient assessment procedures produced an r = 0.93. Some support for content and concurrent validity was noted. Using an external review by an expert panel, a revised instrument and protocol was used to conduct a formal field test of 100 diabetic patients. Levels of internal consistency similar to the pilot, 0.87 and 0.86, were found. An item analysis produced similar positive results. While some support for concurrent validity of the data was found, little support for discriminant validity was evident. The instrument and assessment procedures need to undergo more extensive and rigorous examination of its psychometric characteristics, particularly stability and predictive validity. DIABETES CARE 4: 468–475, JULY–AUGUST 1981.

The importance of quality instrumentation and standardization of measurement procedures in epidemiology, social and behavioral science, and health education is widely acknowledged. However, this recognition has not, until quite recently, led to concern for systematic, analytical examinations of educational assessment instruments or procedures for diabetic patients. More sophisticated, qualitative analyses of self-care data—cognitive, affective, skill, and behavior—for diabetic patients are needed.[1–7] Qualitative issues such as reliability and validity of data and standardization and periodic monitoring of assessment procedures have received almost no attention in previous reports.[8–20] With one exception, the Rand Report (1979), psychometric analyses of the quality of instrumentation or measurement procedures for educational assessment of diabetic patients have not been reported.[4]

This article describes the procedures and analytical methods used to develop an instrument and assessment procedures for diabetic patients. Statistical analyses and methods absent in the diabetes patient education literature but common to psychological and educational measurement were employed.[21–36] This project was conducted over a 2-year period (April 1978 to April 1980) to meet the service needs of the Patient Education Program of the 40-bed Diabetes Hospital (DH) and the research interests of the Diabetes Research Training Center (DRTC) at the Medical Center, University of Alabama in Birmingham. Approximately 1000 inpatient and 12,000 outpatient visits have been recorded each year for the last 3 yr at the DH.

METHODOLOGY

Presented in Table 1 is an overview of the methodology employed in this project. After an extensive retrospective record review to evaluate the then existing situation with respect to educational assessment, it was determined that the instrument and assessment methods in use were neither effective nor efficient, and a new approach was necessary.

First, program behavioral objectives in measurable terms were formulated and defined through a nominal group process. Next, the literature was reviewed and other centers of excellence in diabetes education were contacted to obtain instrumentation and procedures already in use. From these, several draft instruments were developed and reviewed by staff and investigators.

After policy and procedures were established, the first pilot test of 56 patients was conducted to determine the following characteristics of the preliminary instrument: valid-

TABLE 1
Outline of procedures for developing patient assessment instruments and process

1. Formulation of program objectives—review of literature
2. Definition of objective in behavioral or performance terms
3. Review of existing and available instruments and record keeping systems
4. Identification of essential cognitive, affective, and psychomotor skills, and descriptive data information needs by internal review
5. Preliminary construction of instrument and agreement on procedures
6. Identification of measurement methods and coding—interviewer training
7. Pilot testing with 30–50 patients to determine the following essential characteristics:
   A. Validity
   B. Reliability
   C. Adequacy of questions
   D. Ease of administration
   E. Degree of standardization
   F. Efficiency—time
8. Repetition of internal review and modification and external review
9. Formal clinical testing of instrument with 100 patients to reexamine 7(A)–(F)
10. Repetition of internal review and revision in order to modify measurement process and instrument for clinical–field application

ity, reliability, adequacy of questions, ease and time of administration, and degree of standardization. The instrument was modified from the pilot result and through an extensive internal and external review process. Finally, a formal clinical test of the revised instrument and procedures with 100 patients was conducted, replicating the procedures and analysis of data performed in the intial pilot test.

Retrospective record review. The patient assessment system and instrumentation at the DH since July 1976 were reviewed and discussed with the patient educators, nurses, and hospital staff physicians in early 1978. It was determined that a formal evaluation of the then existing situation was necessary.

A retrospective review of patient medical records covering patients admitted during a 12-mo period from September 1976 to September 1977 was conducted in the spring/summer of 1978. Although it was hospital policy that each patient should be assessed soon after admission, no quality control mechanisms existed to monitor this dimension of the patient education program. Only 394 (39%) of 996 patients admitted during the 12-mo review period had an educational form assessment in their patient record file. Only 62 (6%) had a discharge educational assessment on file. Data collected from this review were of such poor quality that they could not be used to perform analyses of patient educational problems or a formative program evaluation. An average of 70% missing data per form was found. Because the response format was not uniform and was highly subjective, the record data present were neither organized to facilitate data processing nor easily codable.

Situational analysis. The first step in developing the new approach was to determine program objectives and the extent to which the objectives were measurable. This was done through a nominal group process resulting in a set of measurable objectives. Concurrent with this activity, the literature, instruments from other sites, and the preliminary results of the Rand Report delphi process were reviewed to determine the utility of each in constructing an instrument and patient assessment system compatible with the objectives.[4] From these, 108 questions (items) were derived for possible inclusion in the patient assessment instrument. Each item was rated independently for degree of essentiality by all the patient educators and selected nursing and medical staff.

Following this initial independent review and rating, monthly group process meetings were held to refine and reach consensus on each question. The resulting draft was critiqued by six nationally recognized experts in diabetes care and education as an external content validity check, revised using comments of the external reviewers, and recirculated for internal staff review.

Pilot study. The pilot study was conducted by the patient education staff in the fall of 1978 with 56 diabetic patients, 46 insulin-dependent and 10 non-insulin-dependent. The assessments took an average of 40 min per patient. The cognitive index (test) of the instrument covered four content categories and consisted of 25 true–false–don't know questions. The performance index included 10 observer criteria covering two skill categories, urine testing and insulin administration. The interviewer asked questions on past instruction, knowledge, skill, and daily self-care practices in the following subject areas: foot and skin care, urine testing, insulin administration, and safety measures. Demographic and life-style questions were included.

Four principal indices were derived for each patient: instruction index (extent and recency of instruction in the subject areas noted), cognitive index (aggregate knowledge score), performance index (self-care skill score), and behavior index (current practices score). Dietary information was not included in this initial form, since at the time dietary assessment was the administrative responsibility of the hospital nutrition staff.

Program and quality control issues such as consistency in scoring, standardization of the interview process, ease of administration, time, and patient comprehension were determined. Internal consistency (reliability), item analyses of various subsections of the instrument, interrater agreement, and concurrent validity (convergent and discriminant) analyses were performed. No stability measurements (test-retest) were possible at this time because of inadequate resources to conduct followup assessments on patients.

Instrument analysis. The reliability (internal consistency) of the cognitive and performance index scores derived from the pilot study was determined by application of a Kuder-Richardson analysis.[22,25] As indicated in Table 2, the reliability coefficients for the cognitive and performance indexes were 0.89 and 0.85, respectively. Reliability coeffi-

TABLE 2
Reliability and standard error of measurement of pilot test data for cognitive and performance indexes

Cognitive index	Performance index
Items (K) = 25	Items (K) = 10
Mean (M) = 15.69	Mean (M) = 7.11
Variance (S^2) = 39.06	Variance (S^2) = 8.64
S = 6.25	S = 2.94
KR_{21} = 0.89	KR_{21} = 0.85
S_m = 2.07	S_m = 1.14
N = 56	N = 46
Formulae: $KR_{21} = \frac{K}{K-1}\left[1 - \left(\frac{M(K-M)}{KS^2}\right)\right]$	$S_m = S\sqrt{1-r}$

cients in the range of 0.70 or above are generally considered to be a minimum standard. The obtained coefficients of > 0.80 suggest a satisfactory level of instrument internal consistency for this setting and group of patients.[22–27]

An item analysis was performed to determine the difficulty of the various questions (percent correct) and their discrimination function (item-to-total correlation). Item analysis data are presented in Tables 3 and 4 for the cognitive and performance indexes. As indicated in column 1 of Tables 3 and 4, with the exception of three items, the item difficulties (percent correct) for these items fell within the generally acceptable range defined in educational measurement of 30–90%.

Item discrimination coefficients were computed for each item to determine the relationship between a patient's score in one item and his/her overall score for the total index. An item with a high positive discrimination coefficient indicates that the item was consistently answered or performed correctly by patients "more knowledgeable or skilled." Item discrimination coefficients (column 2) of Tables 3 and 4 indi-

TABLE 3
Item analysis for pilot trial cognitive data

Category	Item	Item difficulty (%)	Item discrimination
Foot/skin care	Circulation	86	0.35
	Heating pads/hot water	73	0.56
	Corns & callouses	65	0.49
	Wound treatment	52	0.60
Urine testing	Presence of acetone	62	0.64
	Routine test time	50	0.61
	Presence of sugar	60	0.41
	More frequent testing	85	0.58
	Use of color chart	71	0.65
	Use of second void	75	0.57
Insulin	Effect of insulin	83	0.53
	Duration of effect	62	0.62
	Type of syringe	88	0.69
	Disposable syringes	77	0.58
	Adjustment of dosage	54	0.68
	Shelf-life of insulin	60	0.63
	Site rotation	75	0.71
Safety measures	Cause of ketoacidosis	46	0.30
	Symptoms of ketoacidosis	36	0.48
	Self-treatment of hypoglycemia	58	0.40
	Treatment of ketoacidosis	67	0.56
	Use of glucagon	27	0.46
	Need for insulin adjustment	42	0.53
	Regulation during illness	42	0.46
	Need for dietary adjustment	73	0.56

TABLE 4
Item analysis for pilot trial performance data

Category	Item	Item difficulty (%)	Item discrimination
Urine testing	Proper handling	69	0.77
	Proper timing	50	0.76
	Accurate reading	69	0.86
	Records results	56	0.78
	Correct interpretation	68	0.72
Insulin	Proper preparation of materials	88	0.46
	Selection of injection sites	88	0.76
	Preparation of site	94	0.64
	Accurately draws up insulin	75	0.65
	Safe injection	94	0.64

cate that all the items discriminated positively, exceeding the minimum 0.20 standard of acceptability in educational measurement for item analysis. Thirty-three of the thirty-five items exceeded a more rigorous standard of ≥ 0.40.[24,27] In the interpretation of item difficulty and item discrimination data, and their use in the modification of an instrument, it is important to consider both for any given question and question saliency before revising or discarding it.

Validity. The issue of content validity of the instrument was addressed, in part, in the preliminary development steps by the internal and external reviews of staff and experts at the DH and in the field.

A preliminary and limited examination of the degree of convergent and discriminant validity of the instrument was performed. A simplified matrix comparing different traits measured by different methods was used to examine their statistical association. The convergent (concurrent) validity of an instrument is examined by calculating the strength of the association between a characteristic, for example, urine testing knowledge, and a criterion variable, urine testing skill or routine behavior for the same individual at one point in time. In addition to the determination of evidence of convergent validity, the multi-trait multi-method approach can be used to confirm that a particular method of measurement is not highly correlated with another method of measurement from which it should differ (discriminant validity). Of the two types of validity, convergent and discriminant, discriminant is the stronger test of instrument validity.[28,29]

Significant validity coefficients (> 0.01) were found between the total knowledge and performance index scores ($r = 0.56$) and within subject areas, foot instruction and foot behavior ($r = 0.44$), insulin instruction and insulin behavior ($r = 0.50$), and urine instruction and urine testing behavior ($r = 0.34$). While the results of this preliminary examination provided some evidence of convergent validity for selected traits, no evidence in support of the discriminant validity of the variables assessed was found.[25]

Some support for the construct validity of the scale was found in that significantly higher scores on the cognitive, self-reported behavior, and the performance indexes were noted for patients with higher levels of education, with more extensive exposure to diabetes education programs, and who perceived little need for additional instruction.

Standardization of assessment procedures. Another quality control concern addressed by this project was the standardization of assessment procedures. To control for variability of assessment conditions, a standardized kit of materials and equipment for urine testing and insulin administration was used. To establish the level of interrater agreement, an observational trial (trial 1) involv-

TABLE 5
Percent and level of significance of interrater agreement and rater improvement by category and trial

Category	1st trial	2nd trial	Percent gain(+)
I. Demographic	97.3*	99.0	1.8
II. Life-style	89.7*	99.0	10.4
III. Instructional/behavior index			
A. Foot/skin care	93.3	100.0	7.2
B. Urine testing	86.3	98.7	14.1
C. Insulin	89.0	99.3	11.6
D. Reactions/ketoacidosis	94.7	98.7	4.2
Subtotal	90.8*	99.2	9.2
IV. Cognitive index			
A. Foot/skin care	91.3	100.0	9.5
B. Urine testing	96.0	96.7	0.0
C. Insulin	98.0	100.0	2.0
D. Reactions/ketoacidosis	96.0	100.0	4.2
Subtotal	95.3*	99.2	4.2
V. Performance index			
A. Urine testing	73.0	100.0	37.0
B. Insulin prep.—injection	87.0	96.3	10.7
Subtotal	80.0†	98.2	22.3
VI. Total	90.9	99.0	8.9‡

* Kappa analysis (K) of category I + II + III + IV, $Z > 10.0$, $P > 0.001$.
† Intraclass correlation (RI = 0.93), $P > 0.01$ with 7 df.
‡ >0.05 one-tailed t test with 22 df.

ing nine adult insulin-dependent patients and three pairs of interviewer/observers (AB, BC, and CA) was conducted. As seen in Table 5, an analysis of the observed degree of agreement for each of the nine interviews performed by the three pairs of observers in trial 1 was performed using the Kappa (K) statistic for categories I, II, III, and IV. Kappa analysis is applied to nominal data to determine whether an observed level of agreement is beyond the level expected by chance alone. As indicated in Table 5, using the K analysis, significant Z scores in excess of 10.0 were found for all nine paired observations in these categories, > 0.001.[31,32]

Interviewer agreement data presented in Table 5 indicated some inconsistency in one area: the observation of patient skills in category V. An analysis of the level of agreement for category V produced a rater reliability coefficient of 0.93, ≥ 0.001 level. The degree of interrater reliability for category V performances indexes was determined by using an intraclass correlation analysis for interval data. An $R \geqq 0.93$ is considered to be a very high rater reliability coefficient confirming excellent discrimination among subjects and accurate measurement.[32–36]

To reduce the assessor error found in trial 1, a patient assessment training program was conducted. The program focused on specific issues and problems indentified in trial 1, particularly on refining and reaching agreement on criteria to be applied in future assessments (trial 2), and on techniques to reduce extraneous interviewer patient verbal interaction. Following the training program, trial 2 was performed. Data derived from the second field trial showed significant (> 0.05) overall improvement (8.9%) in rater agreement. The striking improvement (22%) for the performance section confirmed an improvement in assessor skill.[30]

RESULTS

Using the experience and results derived from the first year of work, a formal clinical application of the instrument and system was conducted using 100 patients, 73 insulin-dependent and 27 non-insulin-dependent, during June to October 1979. To ensure maximum control over the potential sources of error, three trained interviewers were used in the assessment process. The draft used included the five major categories examined in the pilot test and three additional cognitive sections (general knowledge, diet, and complications) and two new skill areas (urine ketone testing and diet).

Formal clinical test. A comparison of the 100 study patients to over 5000 diabetic patients admitted to the DH in prior years indicated that these patients were comparable by age, sex, and average duration of their disease to the hospital's adult inpatient population without other significant mental or physical disability. The ratio of insulin-dependent to non-insulin-dependent for the hospital population and study sample was comparable, 3:1.

Instrument analysis. Data derived from these assessments were analyzed in the same manner as the pilot test data. As reported in Tables 6 and 7, the item difficulty and item discrimination analyses were performed for the cognitive and performance indexes. Ten items were found to be very easy, $> 90\%$ correctly answered, and six times were found to have low positive discrimination coefficients (< 0.20). The internal consistency coefficients for the cognitive and performance indexes as noted in Table 8 were 0.87 and 0.86, respectively. The data reported in Tables 6, 7, and 8 were highly comparable to the pilot study data in Tables 2, 3, and 4.

Validity. The results of the convergent and discriminant analyses are presented in Table 9. As indicated, with the exception of foot and skin care, four indices were developed for three concept areas: urine testing, insulin use, and diet. For each concept area, an assessment of exposure to instruction, current knowledge, current performance or skill, and regular self-care behaviors was made.

As noted in Table 9, a strong association, 0.60, was ap-

TABLE 6
Item analysis of field trial cognitive data

Category	Item	Item difficulty	Item discrimination
General	Affect on metabolism	82	0.31
	Need for insulin	93	0.27
	Production of insulin	82	0.53
	Effect of insulin	84	0.39
	Source of disease	97	0.40
Foot/skin care	Circulation	94	0.31
	Heating pads/hot water	71	0.47
	Corns and calluses	88	0.20
	Wound treatment	59	0.50
	Shoes and socks	99	−0.02
Urine testing	Presence of acetone	65	0.66
	Presence of sugar	76	0.44
	More frequent testing	88	0.52
	Interpreting sugar results	91	0.54
	Use of second void	70	0.53
	Significance of acetone	58	0.41
Diet	Alcoholic beverages	77	0.43
	Meat exchange	61	0.18
	Vegetable/fruit exchange	51	0.26
	Meal frequency	68	0.27
	Dietary requirements	86	0.51
Insulin	Duration of effect	85	0.44
	Type of syringe	89	0.33
	Disposable syringes	94	0.10
	Adjustment of dosage	62	0.41
	Insulin and urine testing	49	0.19
	Shelf life of insulin	93	0.33
	Insulin reaction	49	0.44
Safety measures	Treatment of hypoglycemia	88	0.44
	Use of glucagon	28	0.49
	Urine testing and illness	84	0.48
	Need for dietary adjustment	71	0.49
	Cause of ketoacidosis	80	0.12
	Symptoms of ketoacidosis	53	0.82
	Treatment of ketoacidosis	87	0.39
Complications	Heat disease	90	0.40
	Blindness	94	0.24
	Pregnancy complications	70	0.43
	Kidney disease	90	0.37

TABLE 7
Item analysis of clinical test performance data

Category	Item	Item difficulty (%)	Item discrimination
Urine glucose testing	Proper preparation	86	0.56
	Proper handling	70	0.54
	Proper timing	29	0.52
	Accurate reading	81	0.62
	Records results	62	0.64
	Correct interpretation	79	0.65
Urine ketone testing	Proper preparation	43	0.71
	Proper handling	38	0.76
	Proper timing	18	0.63
	Accurate reading	41	0.77
	Records results	36	0.76
	Correct interpretation	28	0.66
Insulin	Proper preparation of materials	70	0.48
	Properly draws dose	64	0.50
	Identification of injection sites	94	0.24
	Preparation of site	86	0.37
	Safe injection	59	0.33
	Correct disposal	77	0.20
Diet	Identifies meat exchange	45	0.35
	Identifies starch exchange	27	0.31
	Identifies fruit exchange	87	0.15
	Accurate use of measuring cup	93	0.32
	Accurate use of diet scales	78	0.09
	Quantity in ingredient label	31	0.38
	Identifies sugars on label	91	0.14

parent between the cognitive index and performance index. Data presented in Table 9 highlighted by triangulation examines the degree of convergent validity within each of the four concept areas. Of the four, only the data presented in triangle II for the concept area of urine testing provided support for convergent validity. While several coefficients within other triangles were statistically significant, no other consistent pattern was apparent. Data in triangle I, for example, suggest a very weak association between foot care instruction and knowledge and daily foot and skin care inspection behavior. In triangle III, the data indicate a weak association between insulin instruction, knowledge, and skill. Data in triangle IV present a mixed message in that there was a nonsignificant association between diet instruction and knowledge, yet an apparent strong association between diet instruction and diet planning skill.

In examining the discriminant validity of these data, the only concept area that suggested some degree of discriminant validity was urine testing, triangle II. The blocked coefficient (0.65) was derived for the same concept using two different methods, a urine testing knowledge index and urine testing (interview) performance (skill observation). Using the same method, the derived coefficient, blocked 0.65, was of a considerably greater magnitude than the blocked coefficients, 0.26 and 0.27, for the other concept areas, insulin performance and diet performance.

DISCUSSION

The quality control issues addressed by this project represent one of the first reported attempts to explore assessment problems common to practically all sites at which health care education for diabetes is available. With few exceptions, all programs are required to assess and record information about patients. As indicated, quality control over patient record data necessitates continuous surveillance. A number of quality control procedures can be introduced to promote accuracy and objectivity of the data collected. In performing patient educational diagnostic procedures, a far greater emphasis on data validity and reliability needs to occur than is currently reported. Greater concern, not only to develop instruments of

TABLE 8
Reliability and standard error of measurement of clinical test data for cognitive and performance indexes

Cognitive index	Performance index
Items (K)= 35	Items (K) = 23
Mean (M) = 26.69	Mean (M) = 13.92
Variance (S^2) = 40.13	Variance (S^2)=32.41
S = 6.33	S = 5.69
KR_{21} = 0.87	KR_{21} = 0.86
S_m = 2.28	S_m = 2.12
N =100	N = 73
Formulae: $KR_{21} = \frac{K}{K-1}\left[1 - \left(\frac{M(K-M)}{KS^2}\right)\right]$	$S_m = S\sqrt{1-r}$

TABLE 9
Validity coefficients for clinical test data

	Cog. ind.	Perf. ind.	Ft-S inst.	Ft-S know.	Ft-S beh.	Urine inst.	Urine know.	Urine perf.	Urine beh.	Insulin inst.	Insulin know.	Insulin perf.	Insulin beh.	Diet. inst.	Diet. know.	Diet. perf.
Perf-I	0.60															
FSI	0.33	0.17														
FSK	0.65	0.29	0.18	I												
FSB	0.26	0.05	0.20	0.15												
UI	0.34	0.41	0.52	0.23	0.13											
UK	0.80	0.64	0.32	0.42	0.28	(0.42)										
UP	0.56	0.82	0.16	0.30	0.01	0.39	[0.65]	II								
UB	0.36	0.53	0.17	0.18	0.14	0.47	0.52	0.59								
II	0.34	0.32	0.44	0.27	0.16	0.66	0.37	0.29	0.31							
IK	0.71	0.34	0.14	0.25	0.17	(0.15)	0.43	0.44	0.15	0.23						
IP	0.31	0.67	0.02	0.07	0.02	0.32	[0.26]	0.45	0.23	0.15	0.21	III				
IB	0.24	0.32	0.08	0.17	0.10	0.14	0.21	0.30	0.19	0.40	0.24	0.04				
DI	0.42	0.20	0.39	0.30	0.18	0.45	0.38	0.14	0.24	0.53	0.20	−0.07	0.09			
DK	0.57	0.40	0.23	0.34	−0.08	(0.10)	0.24	0.31	0.07	0.09	0.22	0.22	0.10	0.17	IV	
DP	0.46	0.50	0.04	0.13	−0.21	0.09	[0.27]	0.27	0.13	0.10	0.19	0.21	−0.03	0.27	0.60	
DB	0.13	0.17	0.03	0.03	0.14	0.11	0.19	0.14	0.12	0.11	0.07	0.14	0.06	−0.07	0.11	0.12

Level of sig. at 0.05 ($r \geq 0.20$ with 98 df).
Level of sig. at ($r \geq 0.26$ with 98 df).

high quality, but also to train and evaluate staff adequately in the use of the instruments and assessment procedures needs to become apparent in the literature.

While more work is needed in this area, particularly in strengthening validity, these data suggest that the new instrument and procedures may have the following benefits to an existing program: (1) patients can be efficiently and comprehensively assessed; (2) data collected would exhibit a high level of reliability; (3) data could be routinely processed and aggregated for analysis; and (4) a formative evaluation of the impact of the patient education program through patient baseline and follow-up assessments using the standardized procedures described could be performed.

Because this research was perceived as an initial effort into a relatively unexplored area, the instrument and procedures derived from this developmental study need to be adapted and applied in a variety of settings in a standardized fashion, and to undergo continuous, rigorous examination. Additional analyses of the predictive validity and stability of the traits assessed in this project are needed. Basic patient assessment research is needed with particular emphasis on predictors of health outcomes, control, health services utilization, and incidence of diabetic sequelae.

ACKNOWLEDGMENTS: This research was supported by the Diabetes Trust Fund, Diabetes Research and Education Hospital, the Medical Center, University of Alabama in Birmingham. The authors would like to acknowledge the support of Buris Boshell, M.D., Director, Diabetes Research and Education Hospital, Unversity of Alabama in Birmingham, and his staff, and the assistance of Sharon Jones, M.P.H., and Shirley Boyd, M.A. Parts of this paper have been presented at the Thirty-Ninth Annual Meeting of the American Diabetes Association, Los Angeles, California, June 1979; the 107th APHA Annual Meeting, New York, November 1979; and the Third National Symposium on Patient Education, March 1980, Baltimore, Maryland.

From the Division of Health Education-Health Behavior, Department of Public Health, School of Medicine–The Medical Center, University of Alabama in Birmingham, Birmingham, Alabama 35294.

Address requests for reprints and instruments to Richard A. Windsor at the above address.

REFERENCES

[1] Gordis, L.: Assuring the quality of questionnaire data in epidemiologic research. Am. J. Epidemiol. *1*: 21–24, 1979.

[2] Demlo, L. K., Campbell, P. M., and Brown, S. S.: Reliability of information abstracted from patient's medical records. Med. Care *16*(12): 995–1005, 1978.

[3] Rossi, P. H., Freeman, H. E., and Wright, S.R.: Evaluation: A Systematic Approach. Beverly Hills, Sage Publications, 1979.

[4] Marquis, K. H., and Ware, J. E., Jr.: Measures of Diabetic Pa-

tient Knowledge, Attitudes and Behavior Regarding Self-Care: Summary Report. Rand, Santa Monica, Calif., August 1979.

[5] Gartseff, G., Windsor, R. A., and Jones, S.: Development of an educational needs assessment for diabetic patients. Presented at the 39th Annual Meeting of the American Diabetes Association, Los Angeles, Calif., June 1979. Abstract in Diabetes *28* (Suppl. 2): 350, 1979.

[6] National Conference on Behavioral and Psychosocial Issues in Diabetes (NCBPID): Summary of Proceedings. Diabetes Care *3*: 379–81, 1980.

[7] Windsor, R. A.: Quality assurance in patient educational diagnostic assessments of diabetic patients. Presented at 107th APHA Annual Meeting, New York, N.Y., November 1979, and at the Third National Symposium on Patient Education, sponsored by the Johns Hopkins University, School of Hygiene and Public Health, Baltimore, Md., March 1980.

[8] Beaser, S. B.: Teaching the diabetic patient. Diabetes *5*: 146–49, 1956.

[9] Etzwiler, D. D.: What the juvenile diabetic knows about his disease. Pediatrics *29*: 135–41, 1962

[10] Etzwiler, D. D., and Sines, L. K.: Juvenile diabetes and its management: family, social, and academic implications. JAMA *81*: 304–308, 1962.

[11] Bolt, W. M. and Miller, E. C.: Education of the diabetic patient—physician's dilemma. N.C. Med. J. *28*: 1–6, 1967.

[12] Etzwiler, D. D.: Who's teaching the diabetic? Diabetes *16*: 111–17, 1967.

[13] Fuchsberg, R. R.: The diabetes supplement of the National Health Survey, II. Methods and techniques. J. Am. Diet. Assoc. *52*: 121–24, 1968.

[14] McDonald, G. W.: The diabetes supplement of the National Health Survey, I. Introduction and overview. J. Am. Diet. Assoc, *52*: 118–20, 1968.

[15] Collier, B. N., and Etzwiler, D. D.: Comparative study of diabetes knowledge among juvenile diabetics and their parents. Diabetes *20*: 51–57, 1971.

[16] Singleton, M.: An analysis of the teaching of diabetic patients. J. Med. Assoc. Ga. *60*: 4–10, 1971.

[17] Nickerson, D.: Teaching the hospitalized diabetic. Am. J. Nurs. *72*: 935–38, 1972.

[18] Backscheider, J. E.: Self-care requirements, self-care capabilities, and nursing systems in the diabetic nurse management clinic. Am. J. Public Health *64*: 1138–46, 1974.

[19] Burke, E. L.: Assessment. *In* Nursing Management of Diabetes Melltius. Guthrie, D. W., and Guthrie, R. A., Eds. St. Louis, C. V. Mosby Co., 1977.

[20] Miller, L. V., Goldstein, J., and Nicolaisen, G.: Evaluation of patients' knowledge of diabetes self-care. Diabetes Care *1*: 275–80, 1978.

[21] American Psychological Association: Standards for Educational and Psychological Tests. Washington, D.C., American Psychol. Assoc., 1974.

[22] Richardson, M., and Kuder, G.: The calculation of test reliability coefficients based on the method of rationale equivalence. J. Educ. Psychol. *30*: 681–87, 1939.

[23] Guilford, J. P.: Psychometric Methods. New York, McGraw-Hill, 1954.

[24] Diederich, P.: Short-Cut Statistics for Teacher-Made Tests. Princeton, N. J., Educational Testing Service, 1960.

[25] Thorndike, R. L., and Hagen, E. P.: Measurement and Evaluation in Psychology and Education. 4th edit. New York, John Wiley & Sons, 1977.

[26] Carmines, E. G., and Zeller, R.A.: Reliability and Validity Assessment. Series: Quantitative Applications in the Social Sciences, Sullivan, J., Ed. Beverly Hills, Sage Publications, 1979.

[27] Ebel, R. L.: Essentials of Educational Measurement, 3rd edit. Englewood Cliffs, N. J., Prentice-Hall, 1979.

[28] Campbell, D. T., and Fiske, D. W.: Convergent and discriminant validation by the multi-trait–multi-method matrix. Psychol. Bull. *56*: 81–105, 1959.

[29] Shortell, S. M., and Richardson, W. C.: Health Program Evaluation. St. Louis, C. V. Mosby, Chap. 4, 1978.

[30] Windsor, R.: Improving hospital staff skills in patient educational assessment: a case study in diabetes. Patient Counsel. Health Educ. *3*(1), March/April 1981.

[31] Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Meas. *20*: 37–46, 1960.

[32] Fleiss, J. L.: The measurement and control of misclassification error. *In* Statistical Methods for Rates and Proportions. New York, John Wiley & Sons, Chap. 12, 1973.

[33] Haggard, E.: Intraclass Correlation and the Analysis of Variance. New York, Holt, Rinehart & Winston, 1958.

[34] Fleiss, J. L.: Estimating the reliability of interview data. Psychometrika *35*(2): 143–62, 1970.

[35] Burdock, E. I., Fleiss, J. L., and Hardesty, A. S.: A new view of inter-observer agreement. Personnel Psychol. *16*: 373–84, 1963.

[36] Bartko, J. J.: The intraclass correlation coefficient as a measure of reliability. Psychol. Rep. *19*: 3–11, 1966.