The Validation of a Diabetes Patient Knowledge Test

GEORGE E. HESS, M.A., AND WAYNE K. DAVIS, Ph.D.

Diabetes knowledge tests, used in conjunction with measures of patient attitudes and behaviors, can provide a useful basis for assessing educational needs and designing appropriate instructional experiences. Accurate decisions require instruments that measure patient knowledge of diabetes and its management with high reliability and validity. Data obtained from more than 950 administrations of two parallel forms of a Diabetes Patient Knowledge Test have provided documentation of patient knowledge levels, insight into the effectiveness of educational programs, and support for ongoing program revisions. These data have also allowed study of the psychometric properties of the test instruments, including factor structure, reliability, and validity. Each test form has an overall reliability of 0.89 and the forms are of equal difficulty. Five subcomponents (factors) labeled "Carbohydrates," "Blood Sugar," "Basics," "Food Exchanges," and "Insulin Administration" are measured in the tests. Evidence of content, construct, concurrent, and discriminant validity has been demonstrated. DIABETES CARE 6: 591–596, NOVEMBER–DECEMBER 1983.

The goal of the project described in this article was to develop a knowledge test that would assist health providers in determining the educational needs of patients with diabetes. Following instrument development, it also has been adopted for use in measuring the outcomes of educational programs and as an evaluation instrument in research and evaluation studies. The purpose of this article is to describe the test development process and to document the statistical and psychometric properties of the instrument as determined to date. A secondary purpose is to describe the test validation process to serve as a model for others interested in developing and validating similar tests.

The University of Michigan established, with federal grant support, the Michigan Diabetes Research and Training Center commencing 1 September 1977. Within the first 4 mo of the operation of diabetes clinics within the Center, the need for an easily administered, highly reliable, and valid patient needs-assessment instrument became apparent. At the same time, The Michigan Department of Public Health provided supplemental funding for the development of a needs-assessment instrument subsequently called the Diabetes Educational Profile (DEP). The DEP includes patient-completed instruments that collect information about the patient, his/her living style, family situation, and health regimen. An additional section is used by the nurse to evaluate the performance of the patient in such skills as insulin self-injection and urine testing. One portion of the Diabetes Educational Profile is a knowledge test, which was designed to measure the patient's store of information regarding diabetes and its treatment. In this article we will describe the development and psychometric properties of this knowledge test.

METHODS

Test development. The technical expertise for the development of the knowledge test came from health professionals and educational evaluation experts. The physicians, nurses, and dietitians within the Michigan Diabetes Research and Training Center provided the content expertise from which the educators developed a test outline. A detailed outline, which included the areas of content the test should cover, was developed. Once consensus was reached as to the content outline, the content areas were refined further and written in the form of instructional objectives. These objectives formed the basis of the inpatient educational classes held in the Diabetes Center Unit. Each of the objectives was then matched to test items that measured the objective. In all, over 150 test items were collected from a wide variety of sources including test instruments used within the Diabetes Center

Unit of University Hospital and other medical centers. After a pilot use of the test on diabetes patients who were hospitalized in University Hospital, item analysis was performed and the resulting data used to reduce the number of items on the test from 150 to 38. Items were removed (1) when they did not discriminate the most knowledgeable patients from the least knowledgeable or (2) when the answers did not contribute to the understanding of patient learning needs. Pairs of parallel items were identified, based on item content and their response distribution with pilot studies. These parallel items were then randomly assigned to create two alternate forms of equal difficulty and comparable content. Each form contained 38 items. Each test form was then used both as a pretest and a posttest* to allow confirmation of the comparability of the two forms.

Psychometric properties. To determine if the responses by the patients formed patterns similar to the content patterns identified in the initial test outline, factor analysis was performed on Form B of the test. Form B was selected as there were more pretests of Form B than Form A. Only pretest administrations of the instrument were used in the factor analytic procedure to avoid the possible influence of instruction received during treatment on the pattern of responses. After the factor analysis of Form B, the items in Form A were assigned to the same factors as Form B according to the factor loadings of their "parallel" items in Form B. After computation of the factor structure, factor reliability estimates were made for *both* test forms using Cronbach's coefficient alpha.[1] This statistic indicates the internal consistency of item responses within each test factor. In computing factor reliabilities (Cronbach's alpha) and item-total score correlations, both pre- and posttest administrations were used. This sample was used since studies relating item discrimination indices to instructional sensitivity[2-4] support use of a sample containing both instructed and noninstructed students.

The patient knowledge test was intended to assess instructional need among literate adults with diabetes and to measure outcomes of educational programs. With these objectives in mind, evidence of content validity, instructional sensitivity, construct validity, concurrent validity, and discriminant validity was gathered.

Generally, content validity reflects the extent to which a test measures a specific domain of content. In this instance, content validity referred to the relationship between knowledge necessary to manage diabetes and test content. The basis for content validity in this study was established by the process used to develop the test. The objectives and class outlines of the inpatient diabetes classes provided the basis for item development. These objectives were compared with two studies that identified cognitive factors related to patient self-management of diabetes mellitus, the Diabetes Educa-

*Pre- and posttests can be obtained by writing to Wayne K. Davis, Ph.D., GIIII Towsley Center, Box 57, Office of Educational Resources and Research, University of Michigan Medical School, Ann Arbor, Michigan 48109.

TABLE 1

Factor structure, reliability estimates, and item-factor correlations of two forms of the Diabetes Knowledge Test

Form A Item	Form A Correlation	Form B Item	Form B Correlation
Factor 1: Carbohydrates			
Form A (alpha = 0.78)		Form B (alpha = 0.76)	
22	0.47	22	0.55
23	0.64	23	0.37
24	0.64	24	0.58
25	0.62	25	0.63
26	0.59	26	0.59
28	0.27	28	0.28
Factor 2: Blood Sugar			
Form A (alpha = 0.86)		Form B (alpha = 0.80)	
7	0.61	7	0.50
8	0.50	8	0.46
9	0.39	9	0.57
10	0.63	10	0.55
11	0.67	11	0.45
12	0.63	12	0.55
13	0.66	13	0.50
14	0.54	14	0.23
15	0.53	15	0.57
16	0.57	16	0.49
Factor 3: Basics			
Form A (alpha = 0.65)		Form B (alpha = 0.56)	
1	0.31	1	0.19
2	0.21	2	0.39
3	0.28	3	0.30
4	0.31	4	0.17
5	0.37	5	0.30
6	0.40	6	0.14
35	0.37	34	0.26
36	0.24	35	0.37
37	0.38	36	0.14
38	0.29	37	0.14
		38	0.22
Factor 4: Food Exchanges			
Form A (alpha = 0.72)		Form B (alpha = 0.64)	
27	0.20	27	0.43
29	0.59	29	0.46
30	0.53	30	0.44
31	0.55	31	0.39
32	0.49	32	0.29
33	0.44	33	0.22
34	0.24		
Factor 5: Insulin			
Form A (alpha = 0.63)		Form B (alpha = 0.59)	
17	0.38	17	0.39
18	0.40	18	0.42
19	0.35	19	0.33
20	0.39	20	0.47
21	0.38	21	0.16

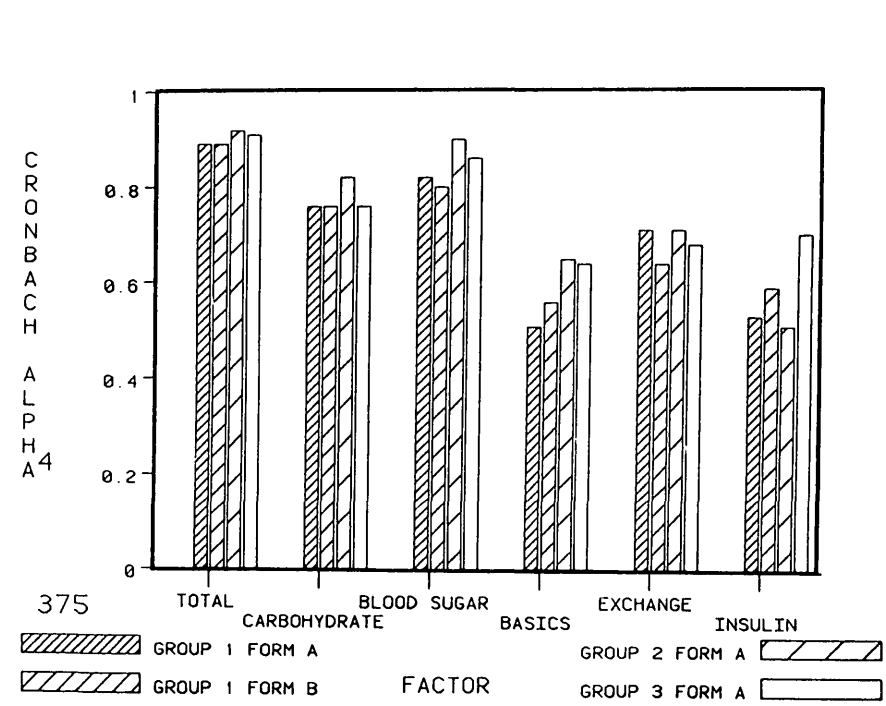

FIG. 1. *Diabetes Knowledge Test estimated reliability.*

tional Profile project,[5-7] and a Rand Corporation study.[8] The content included in the final version of the test was reviewed by a panel of nationally recognized consultants consisting of a diabetologist, a dietitian, and a nurse educator. This review, as well as the previously mentioned comparisons with other published studies, confirmed the appropriateness of test content and thereby established its content validity.

RESULTS

Psychometric properties. A principal axis factor analysis of 210 cases resulted in a five-factor solution accounting for 37% of the variance in test response.* There was almost perfect correspondence between the knowledge categories in the prospective test outline, and the results of the factor analysis. The factors identified were interrelated. Correlations among them range from −0.48 to +0.45.

The first factor, "Carbohydrates," contained six items related to the identification of foods high in carbohydrate content and to sick-day carbohydrate management. A second factor, "Blood Sugar," contained 10 items concerning signs, causes, and management of abnormal glycemic status. The third factor, "Basics," included 11 items basic to treatment goals, urine testing, complications, and control status. Six items were included in the fourth factor, "Food Exchanges." The last factor, "Insulin Administration," contained five items regarding insulin management and administration.

Internal consistency reliability estimates (Cronbach's alpha) and item-total correlations for each factor in both test forms are tabulated in Table 1. The overall reliability estimate for each test form was 0.89. As can be seen in this table, the reliabilities and item-factor correlations are very similar for both test forms, indicating equivalence of test structure.

*A simultaneous promax rotation of the factor structure further delineated the association of items within these factors.

The stability of factor reliability when the test was used with different patient samples is graphed in Figure 1. Group 1 was composed of individuals admitted to University Hospital for care of their diabetes. Two hundred eighty-nine administrations of Form A and 258 administrations of Form B were obtained from this group. Each form was used as both a pretest and a posttest in this sample. The common link among patients in group 2 is that all were under the care of diabetologists. One hundred eighty-one administrations of Form A were obtained from this group. Group 3, a random sample of diabetic patients from eight Michigan communities, provided 228 administrations of Form A. Only pretests were obtained from the last two groups of patients.

The equivalence of test difficulty of the two forms was investigated using 92 scores on Form A and 199 scores on Form B obtained from independent subsamples of group 1 before systematic patient education. Analysis of variance on total test score and on each of the five subscores was used to determine equivalence. The resulting F statistics ranged from 0.06 to 1.9. None was significant and the hypothesis of test equivalence could not be rejected.

To test the similarity of factor structure between the two forms of the test, factor reliabilities were examined. An examination of Figure 1 indicates that the factor reliabilities for Form A (using parallel items in the factor structure) are at least as high as those for Form B (the form on which the factor analysis was based). The fact that the reliabilities of the factors of Form A and Form B were similar confirmed the initial factor solution for both forms.

Validity. Content validity was established by the process of test construction (see METHODS). Another measure of instrument validity was reflected in its ability to detect knowledge gains subsequent to instruction. This "instructional sensitivity" was tested using a sample which included pre- and postinstruction administrations. Discrimination in-

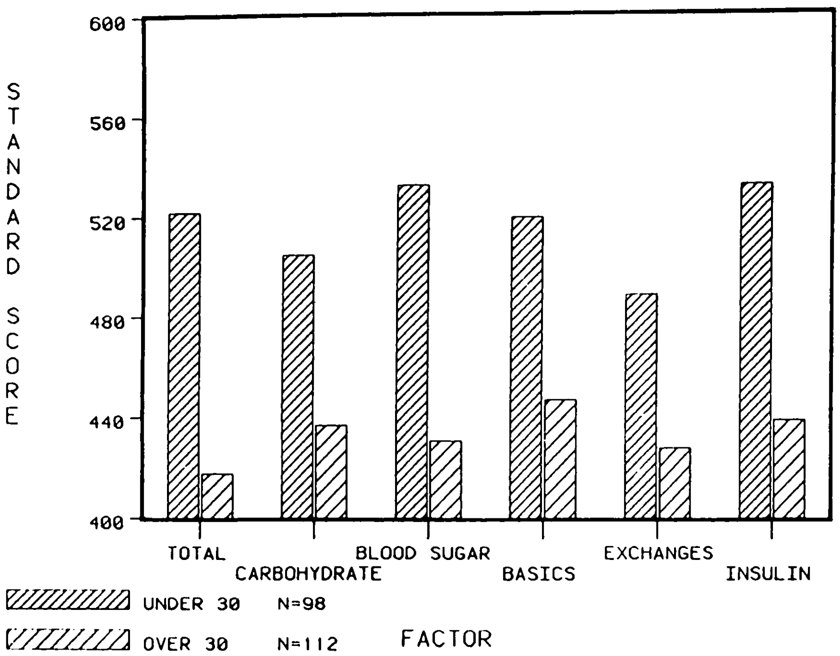

FIG. 2. Diabetes Knowledge Test by age category at onset.

dices in the form of correlations between item responses and factor scores are included in Table 1. Since each sample included both pre- and postinstruction data, the magnitude of the correlations was considered a measure of the utility of the items in assessing instruction. Since the correlations were high, the utility of the instrument was also high. Pre-post differences in factor scores obtained from group 1 ranged from 0.44 to 0.99 standard deviation. These substantial gains supported instructional sensitivity, the ability of the instrument to detect increases in knowledge due to instruction.

Construct validity refers to the relationship of content measured by the test that is understood or inferred from the interrelationships among test items. Based on their clinical experience and their understanding of how patients organize their knowledge of diabetes, content experts identified topic areas that comprised the domain of knowledge of diabetes the test was developed to assess. These topics were reflected in the test outline and described in METHODS. As described previously in RESULTS, statistical analysis revealed five test factors: Carbohydrates, Blood Sugar, Basics, Food Exchanges, and Insulin Administration. The relationships among these factors and the topics of the original test outline were used as a measure of the test's construct validity. Essentially, this procedure assessed whether the structure of knowledge reflected by test performance matched the structure developed by the experts in the content outline. The present test's construct validity was excellent. The only deviation was that items identified in the original outline as (1) disease process, (2) urine testing, and (3) complications had similar response patterns and thus all were grouped together in the "Basics" factor in the factor analysis.

Concurrent validity is a measure of the extent to which the test may be used to estimate a patient's present standing on some criterion or dependent variable. Figure 2 graphs mean pretest scores for two samples (one group was diagnosed before age 30, the other after age 30). Age of diagnosis was used as a surrogate measure for type of diabetes. It was assumed that the first subgroup (the type I's) would be more knowledgeable about the content measured in the test than the type II's (those diagnosed after age 30). This hypothesis was supported. Clearly, the test was sensitive to differences in knowledge levels between these two groups since differences between mean scores ranged from 0.6 to more than 1 standard deviation.

Additional support for concurrent validity is shown in Figure 3. Patients who were discharged taking insulin scored higher on all factors than those who were controlled with oral medications or diet alone. Patients treated with insulin should be expected to score higher if the test is valid. The reasoning is that the majority of these patients have been receiving insulin for some time, and since insulin cannot be administered without instruction, which includes content measured by the test, pretest performance by these patients should be significantly better than the performance of patients not taking insulin.

Discriminant validity is estimated by the extent to which the test does *not* measure irrelevant factors. It is particularly important that a test is not extremely sensitive to the general educational level of the respondents. If high correlations between test scores and educational level did exist, the test would only be measuring general education, not the specific content of diabetes. In this case, the correlations between test score and educational level ranged from 0.07 for the exchange factor to 0.20 for the Basics factor. Correlations of this low magnitude indicated the test measured more than general education level. Moreover, the lower correlations

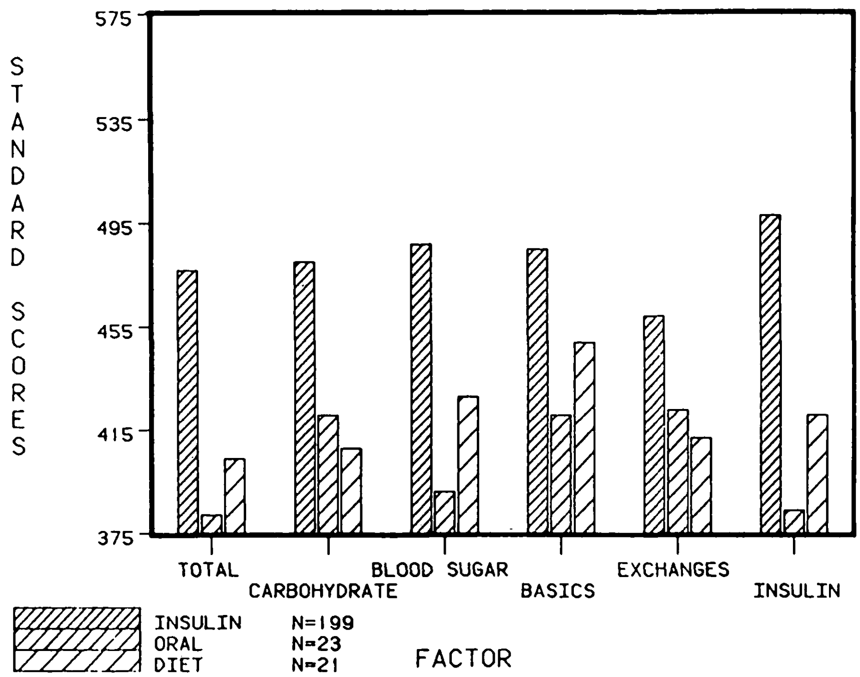

FIG. 3. *Diabetes Knowledge Test by subsequent treatment.*

were associated with more specialized knowledge, and the higher correlations with knowledge of more general aspects of diabetes. Since knowledge of general aspects of diabetes might reasonably be expected to be associated with general knowledge, these values support discriminant validity.

DISCUSSION

This article has described the development of a Diabetes Patient Knowledge Test and subsequent analyses to determine its factor structure, reliability, and validity. Results have indicated that the five-factor test (both Form A and Form B) provided useful information about patient knowledge of diet, insulin, and other aspects of management. The test also provided the basis for the development of educational programs designed to meet the needs of the individual patient. The data have demonstrated that the test was highly reliable and that this reliability was consistent across varying groups of patients. In terms of validity, evidence of content, construct, and concurrent and discriminant validity have all been demonstrated. It should be noted that it was impossible to demonstrate validity through correlations of the results of this test with any known measure of diabetes knowledge as no objective data were found that demonstrate the validity of another diabetes knowledge test.

In addition to the data discussed in this article, several other indicators of test validity include demonstrated differences in scores among patients divided by age of onset, duration of disease, and age at time of test administration. These data all support the concurrent validity of the test.

The test, although quite satisfactory in its current format, does have deficiencies. The reading level of the test is at approximately the eighth or ninth grade level, exceeding the skills of many patients with diabetes. To address this problem, a draft of a test written at a lower reading level, which measures the same content, is being developed.

A second deficiency is the high correlations among factors. Although it is true that patients knowledgeable about food exchanges are probably also knowledgeable about their medication, more discrete and less overlapping knowledge factors would be desirable for diagnosing specific educational needs. It is unlikely that discrete scales will ever be developed as diabetes subject matter is interrelated. Therefore, it is not surprising that the test scores demonstrated these relationships.

The third deficiency in the tests is the result of a compromise. It would be desirable to have a longer test with more items that would more accurately pinpoint the specific needs of the patient. However, the time available in clinical settings and concern for patients' attention spans demanded that the test be short enough to administer in a 20-min period.

Of course, a knowledge test is only a very small component of a complete needs-assessment of a diabetic patient. At the MDRTC, psychological problems, social support, living environment, regimen, and barriers to compliance in addition to knowledge are all assessed. Others are encouraged to collect and use similar information.

In summary, both forms of the Diabetes Patient Knowledge Test are reliable, valid, and measure content judged by health providers to be important for the care of the diabetic patient. The tests provide data upon which educational programs can be based and information useful in assessing the impact of educational interventions with diabetic patients.

ACKNOWLEDGMENTS: The research reported in this article was supported in part by The National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases, Michigan Diabetes Research and Training Center (NIH 2P60AM20572-06), and The Michigan Department of Public Health, Division of Chronic Disease Control.

From the Department of Postgraduate Medicine and Health Professions Education, University of Michigan Medical School, Ann Arbor, Michigan.

Address reprint requests to Wayne K. Davis, Ph.D., GIIII Towsley Center, Box 57, Office of Educational Resources and Research, University of Michigan Medical School, Ann Arbor, Michigan 48109.

REFERENCES

[1] Cronbach, L. J.: Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16:297–334.

[2] Helmstadter, G. C.: Comparison of traditional item analysis selection procedures with those recommended for tests designed to measure achievement following performance-oriented instruction. Presented at the meeting of the American Psychological Association, Hawaii, 1972.

[3] Haladyna, T. M.: Effects of different samples on item and test characteristics of criterion-referenced tests. J. Educ. Measurement 1974; 11:93–100.

[4] Bernknopf, S., and Bashaw, W. L.: An investigation of criterion-referenced tests under different conditions of sample variability and item homogeneity. Presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.

[5] Davis, W., Hull, A., and Boutaugh, M.: Diagnosing the Educational Needs of Patients with Diabetes Mellitus. *In* New Directions for Continuing Education, Assessing Educational Needs for Adults. Pennington, F., Ed. San Francisco, Jossey-Bass Publishers, 1980.

[6] Boutaugh, M., Hull, A., and Davis, W.: An examination of diabetes educational assessment forms. The Diabetes Educator 1982; 7:29–34.

[7] Davis, W., Hull, A., and Boutaugh, M.: Factors affecting the educational diagnosis of diabetic patients. Diabetes Care 1981; 4:275–78.

[8] The Rand Health Insurance Study, Medical Care, XVII:7 Supplement, 1979:1–55.