ISSN 0378-5254 Tydskrif vir Gesinsekologie en Verbruikerswetenskappe, Vol 34, 2006

Development of a Valid and Reliable Test for Higher-Educated Young Adults Measuring Dietary Fibre Food Source and Health-Disease Association Knowledge

Irma Venter

OPSOMMING

Die doel van hierdie studie was om 'n toets te ontwikkel met dieetvesel as die kennisveld ten einde die geldigheid- en betroubaarheidskwessies rakende voedingkennisbepalings te oorkom. Veertig veelkeusige items is opgestel. Dertien items is tydens die beoordeling van die voorkomsgeldigheid uitgegooi. Dit het die veselterminologie-items ingesluit, sowel as die meeste items wat met veseleienskappe, die funksies van vesel en die dieetaanbevelings verband gehou het. Tien bykomende items is daarna opgestel. Die vraelys het grootliks op voedselbronne wat vesel bevat en praktiese keuses daarvan gefokus en tot 'n mindere mate op siektes wat met veselinname verband hou. Dit is gedoen op aanbeveling van die paneel wat die voorkomsgeldigheid van die items beoordeel het en met ondersteuning uit die literatuur. Al die items is volgens die reëls vir itemkonstruksie opgestel. Die vraelys, wat uit 37 items bestaan het, is deur twee groepe hoëronderwysstudente (n=99 en n=87 onderskeidelik) beantwoord. Die veronderstelling was dat die groepe sou verskil wat voedingkennis betref vanweë hulle ingeskrewe programme. Sewentien vrae is ná die item-analise behou. Ses items het onderskeidelik nie die moeilikheidsindeks en die kriteria vir item-tot-totale-korrelasie geslaag nie, sewe items nie die onderskeidingsindeks-kriteria nie en 13 items nie die veranderlike kriteria nie. Die 17-item-toets is as 'n geldige en betroubare kennismeetinstrument bevestig. 'n Betekenisvolle verskil ($p < 0,001$) in kennis is in die verwagte rigting tussen die twee groepe studente aangedui deur van die Mann-Whitney-toets gebruik te maak. Die betroubaarheidskoëffisiënt, bepaal deur die Kuder-Richardson-formule 20 (interne konsekwente metode), was 0,9147. Items is volgens hulle moeilikheidsindekse in die toets gerangskik. Die kennistoets kan in 'n groep of individueel gebruik word en die inligting wat ingewin word, kan van waarde wees vir die ontwikkeling van voedingkennistoetse.

— Ms Irma Venter
Department Food and Agricultural Sciences
Cape Peninsula University of Technology

The input of the item review panel and the education specialists is appreciated. A special word of appreciation to Ms E Hinrichsen for the loading of the data, Mr R van Gijsen for the statistical analysis of the data, and the Cape Peninsula University of Technology for financial support.

INTRODUCTION

Public health nutrition education accentuates fibre consumption to reduce chronic disease risk (Colavito *et al,* 1996). However, fibre intake of Southern African populations has fallen to 20-25 g and 15-20 g daily in rural and urban areas respectively (Walker *et al,* 2001). A greater number of those respondents from the Cancer Control Supplement of the 1987 National Health Interview Survey who indicated that they had made dietary changes for health reasons reported eating more vegetables (44,5%) and fruit (24,9%), than eating less refined grain products (6,5%) (Cotugna *et al,* 1992). An institution based food product development project was initiated to determine the sensory acceptability of bakery products with increased fibre contents due to the belief that fibre has a negative impact on taste (Colavito *et al,* 1996). As part of this project, the fibre knowledge of the target population of young adults, represented by higher education students, had to be determined as nutrition knowledge is a factor linked to eating behaviour (Steenhuis *et al,* 1996; Pirouznia, 2001) that may influence product acceptability.

Reported results on eating behaviour and its relation to nutrition knowledge are, however, inconsistent (Steenhuis *et al,* 1996; Pirouznia, 2001). One of the reasons suggested for the inconsistency is that the nutrition knowledge could have been poorly assessed (Kristal *et al,* 1990; Parmenter & Wardle, 1999). Parmenter *et al* (2000) indicate that the use of measures with little consideration of reliability and validity issues has limited the understanding of nutrition knowledge amongst the United Kingdom (UK) public. Nutrition knowledge questionnaires generally have limitations in the psychometric measures (Axelson & Brinberg, 1992; Steenhuis *et al,* 1996; Parmenter & Wardle, 1999). The consequence of measuring knowledge by means of a questionnaire of unknown validity or reliability is that it is impossible to know whether it actually measures what it claims (i.e. knowledge) (Parmenter & Wardle, 2000). Accurate assessments of the nutrition knowledge-dietary behaviour relationship require use of valid and reliable nutrition knowledge measures (Sapp & Jensen, 1997).

Although Parmenter and Wardle (2000) indicate that a new knowledge measure should only be developed when an instrument cannot be found, they also indicate that investigators often develop their own questionnaires. The advantage provided by this action is that the questions asked could be exactly relevant to the study (Axelson & Brinberg, 1992; Parmenter & Wardle, 2000).

Axelson and Brinberg (1992) emphasise that knowledge will be a good predictor of behaviour if the knowledge measure represented those nutrition aspects that correspond with the dietary behaviour under study.

Parmenter and Wardle (2000) further indicate that a questionnaire developed and validated in one country may not be valid in another due to variations in eating habits and specific dietary recommendations (Parmenter & Wardle, 1999). The decision to develop, rather than obtain, a test was also based on additional factors, such as the knowledge domain selected (i.e. dietary fibre and not general nutrition), the length of the test (i.e. number of questions included should not be too time-consuming to complete), familiarity of examples used for items and answers (i.e. South African), and the inclusion of contemporary fibre nutrition-health issues (i.e. soluble fibre and heart disease). The objective of this study therefore was to develop a valid and reliable test with dietary fibre as the knowledge domain.

METHODOLOGY

Domain and content selection

Dietary fibre formed the content domain. Items that measure the same learning outcome (Gronlund, 1993:97) or deal with the same content (Thorndike *et al,* 1991:239; Parmenter & Wardle, 2000) are usually grouped together. Items that belonged to the same fibre topic were grouped together either as terminology (types of fibre), characteristics, functions, dietary recommendations (guidelines and goals), food sources or diet-disease associations. These topics were used by other researchers who determined fibre knowledge (Resnicow & Reinhardt, 1991; Auld *et al,* 1998). As education stands out as the most important predictor of knowledge scores (Levy *et al,* 1993), the inclusion of questions on terminology and characteristics was seen as justified because the study group would be higher education students.

Test item construction

Selection-type items, which include multiple-choice, true-false and matching formats and exclude supply-type items such as short answer or fill-in and essay formats (Thorndike *et al,* 1991:210; Gronlund, 1993:28), were chosen as the item type. The multiple-choice format was chosen for the type of response because such items are less time-consuming and easier to complete, and scoring and processing simpler (Babbie, 1975:107; Gronlund, 1993:29). Furthermore, this format is considered objective and reliable as the scoring procedure can be stated in advance of testing and is usually clear (i.e. there is only one correct answer) (Nunnally, 1972:155; Gronlund, 1993:46). It is indicated as the most highly regarded and useful selection-type item (Gronlund, 1993:60) and is strongly recommended for use in objective tests (Nunnally, 1972:172). The multiple-choice format was chosen rather than the true-false format to eliminate guessing (Nunnally, 1972:171; Gronlund, 1993:46). Nunnally (1972:160) does not recommend true-false items for general use and indicates that the only way to lower measurement error due to guessing is to make the test long – more than 60 items. Matching items were also not considered. This format is usually utilised when the same alternatives are repeated in several multiple-choice items (Gronlund, 1993:69), which in this case was not apparent. On the other hand, supply-type items take more time to complete and process (Nunnally, 1972:159; Gronlund, 1993:80), the spelling ability of respondents could introduce subjectivity (Gronlund, 1993:79) and response uniformity would not be provided (Babbie, 1975:107).

An item must have at least three answers or distracters to be classified as multiple-choice (Thorndike *et al,* 1991:223). Although three answers are acceptable to decrease guessing (Gronlund, 1993:41), four answers were included. Thorndike *et al* (1991:223) indicate that the typical pattern is to have four or five answers to reduce guessing. Nutrition textbooks and test banks consulted also included four answers. The choice of distracters is important as it permits control of item difficulty (Thorndike *et al,* 1991:229). The use of "none of the above", "all of the above" and other such alternatives were not included as its use is not recommended (Nunnally, 1972:178; Thorndike *et al,* 1991:233). The stems of the items were constructed as questions and incomplete statements (Nunnally, 1972:169; Thorndike *et al,* 1991:223; Gronlund, 1993:41). Comprehension and application type items, which provide the basic means of understanding (Thorndike *et al,* 1991:227; Gronlund, 1993:45), were included.

It was decided that the final test should consist of approximately 20 items as interpretations based on fewer than 10 items are considered tentative (Gronlund, 1993:38), and to ensure that the test is not too long and time-consuming to complete. Forty items were drafted because it is recommended that twice the number of items required should be formulated so that enough items are retained after the item analysis (Huysamen, 1986:46; Parmenter & Wardle, 2000). According to Gronlund (1993:37) and Parmenter and Wardle (2000), writing more items than required makes it possible to discard weak or inappropriate items during the item review. Increasing the number of items may also increase test reliability (Nunnally, 1972:156; Gronlund, 1993:176). Although some items were taken from existing questionnaires and the literature, most were adapted or new items generated.

The rules for item construction were applied in drafting the items. Clarity, precision and relevance of items and avoiding double-barrelled and biased items were some of the requirements that had to be met (Nunnally, 1972:172-181; Babbie, 1975:108-110; Thorndike *et al,* 1991:225-233; Gronlund, 1993:47-60). Some negatively formulated items (n=3) were included, although not recommended for inclusion as they can be misinterpreted (Babbie, 1975:132). According to Gronlund (1993:51), negative wording can be used if it is required for the measurement of an important learning outcome. These negatively worded items followed the construction rules of negative items (Nunnally, 1972:179; Gronlund, 1993:51). The negative wording was emphasised (in bold print) to draw the respondent's attention to it

(Thorndike *et al,* 1991:227). The items were also constructed in a manner to ensure that information given in one did not provide an answer to another (Nunnally, 1972:179; Thorndike *et al,* 1991:240; Gronlund, 1993:37). The distribution of answers to the four alternatives was also positioned randomly (Nunnally, 1972:175; Gronlund, 1993:58) to ensure that each alternative was equally represented as the correct answer (Thorndike *et al,* 1991:241).

Test item review

The items were drafted by a dietician and reviewed for content validity by two registered dieticians with research experience who are actively involved in the field of nutrition training, and two higher education food science lecturers. Dieticians or nutritionists are qualified in nutritional matters to give advice in this regard (Parmenter & Wardle, 2000). Content validity (referred to by Sapp & Jensen (1997) as representative validity) is a concern in knowledge testing. It refers to how well the items represent the content domain measured (Gronlund, 1993:175). Content validity is evaluated by careful logical analysis (Gronlund, 1993:163) based on the professional judgement of subject-matter experts (Thorndike *et al,* 1991:145). As the analysis is rational and judgemental, this process has been referred to as rational or logical validity (Thorndike *et al,* 1991:124). The items were reviewed in terms of accuracy, appropriateness or relevance (Babbie, 1975:109), representativeness of topics covered, suitability and mutual exclusiveness of answers (Babbie, 1975:107), as well as item clarity (Babbie, 1975:108) and format (Huysamen, 1986:40; Nunnally, 1972:29).

The four higher education lecturers, who represent the food science and nutrition fields, and four final-year students, who are knowledgeable about food and nutrition, and who are familiar with the study group they represent, reviewed the drafted items for reasonableness. This appearance of reasonableness is called face validity. Face validity relates to the reasonableness of the test from the respondent's point of view (Thorndike *et al,* 1991:126) – that is, that items may be regarded as either too easy or too difficult (Huysamen, 1986:41).

Face validity was a major concern. The review panel agreed that those items related to fibre terminology, characteristics, functions and recommended dietary intake were possibly too difficult and would not be considered reasonable from the respondent's point of view. To obtain support, the four final-year student panel members each interviewed three students, representing the study group, in small-group discussions. Two students received formal nutrition instruction, while the majority (n=10) did not. These discussions confirmed that the respondents would not be competent to answer most of these drafted items. Gronlund (1993:37) emphasises that the difficulty level of items should match the intended learning outcome and the use to be made of the results. All items related to fibre terminology and most items related to fibre characteristics, functions and recommendations were therefore discarded (n=13). It was decided beforehand that if three or more panel members judged an item to be inappropriate, it would be deleted from the item pool.

Considerations during test item review based on literature In the study by Parmenter and Wardle (1999), the understanding of terms was also removed as these items were judged too scientific and not relevant to behaviour. The review panel in this study also used comments like “too advanced”, “too scientific”, “too technical” and “too physiological” to describe these items, and questioned their relevance. Buttriss (1997) further found that, although more than 95% of UK respondents had heard of the term fibre, only 70% felt confident explaining what the term meant. Regarding the exclusion of the fibre characteristics and functions, a survey determining nutrition knowledge of primary care Canadian physicians indicated that only 39% knew that soluble fibre was the dietary fibre that lowered blood cholesterol (Temple, 1999).

Parmenter *et al* (2000) found that although more than 90% of their sample of the English public was aware of the recommendation to increase fibre intake, indicating that this basic message was being conveyed successfully, 70% did not know that the recommended daily intake of fruit and vegetables was as many as five servings. Just over 50% believed one to three portions to be adequate. In this context, 78% and 52% of urban white and black adult South African samples respectively were aware of eating (more) high-fibre foods, but also respectively indicated that on average three and two-and-a-half servings of fruit and vegetables should be consumed per day, which is far below the recommended five servings (Peltzer, 2004). These results imply that, although a large number of people might be aware of the recommendation, many might not be well informed about it. Harnack *et al* (1997) also indicate that many Americans are confused by the dietary recommendations.

The review panel suggested that the items should focus largely on food sources and practical food choices, and to a small extent on fibre-disease associations. This decision was supported by the study by Peltzer (2001), who found that South African university students seemed to have above-average knowledge on the food sources of fibre (5,06 ± 0,10 out of 10). The 13 items deleted were replaced with 10 items covering fibre food sources and practical food choices. The questionnaire now consisted of 37 items of which most (75 to 80%) were related to food sources and practical food choices.

Some concern, however, still remained regarding the inclusion of these food-based items. Parmenter *et al* (2000) found that the section on fibre in their questionnaire was generally answered well by their sample of the UK public. These findings represented an improvement on those of Buttriss (1997), who found that the UK public was generally poor at identifying foods containing starch and even worse at knowing which foods contained fibre. Although nine of ten respondents in this 1992 sample had heard of the term fibre, 35% were unable to identify even three of six fibre-containing foods. When this question was repeated in 1995, the situation had worsened as 42% of the sample failed to identify half of the fibre-rich foods. Cremer and Kessler (1992) also indicated that Americans appear to lack an understanding of the fibre content in foods. Only one of the four high-fibre foods, bran flakes, was mentioned as such by their re-

spondents. The researchers indicated the wide advertising of high-fibre cereals as the likely reason for this. Despite the high level of fibre awareness found in the United States (US) and Geneva population samples reported by Girois *et al* (2001), a substantial proportion of the respondents in both populations also lacked the ability to rank common foods as high or not high in fibre. In contrast, a relatively high percentage (76.1 to 93.5%) of the respondents could identify the high-fibre choice correctly in five pairs of food items in the 1989 and 1990 Continuing Survey of Food Intake of Individuals (CSFII) and the companion Diet and Health Knowledge Survey (DHKS), but only 58% could identify kidney beans as the correct choice in the lettuce/kidney beans pair (Variyam *et al*, 1996). Furthermore the South African nutrition knowledge study by Peltzer (2004) found that all participating groups had low scores on choosing everyday healthy foods despite reasonable knowledge of sources of nutrients.

Other adjustments made to the items included the following: some changes to grammar and wording; using simpler and more familiar terminology such as vitamins and minerals (replacing micronutrients), starch (replacing carbohydrates), bran (replacing cellulose), blood sugar (replacing blood glucose), etc.; clarifying the term "legumes", which was seen to be an unfamiliar word, by explaining that it includes dried beans, peas and lentils; shortening distracters in which foods were ranked from the highest to the lowest fibre contents from four to three to simplify answering; changing distracters to include either food or beverages or examples from the same food grouping to provide more uniformity; and replacing some distracters with foods more representative of South African eating habits.

The 37 items retained after the content and face validity review formed the knowledge questionnaire. Two education specialists also reviewed these items for reading level, vocabulary, grammar (Gronlund, 1993:36), and item construction rules (Nunnally, 1972:172-181; Babbie, 1975:108-110; Thorndike *et al,* 1991:225-233; Gronlund, 1993:47-60).

Knowledge questionnaire development and testing

A self-administered questionnaire was developed as the respondents could read and write and were familiar with answering questions. Demographic questions were placed at the end of the questionnaire, as advised by Parmenter and Wardle (2000), as respondents might dislike answering them, seeing them as intrusive or threatening. Besides gender and age, the section covered formal nutrition education and other nutrition instruction received, and perceived nutrition knowledge level.

Brief instructions for completion were clearly and prominently indicated (Thorndike *et al,* 1991:240; Parmenter & Wardle, 2000). An introductory comment was provided to explain the purpose of the questionnaire administration. To ensure that the questionnaire did not intimidate respondents, it was not referred to as a "test" (as advised by Parmenter & Wardle, 2000). Participants had to circle the letter (a, b, c or d) preceding the appropriate answer on the questionnaire to indicate their choice. Babbie (1975:112) sees this as a favourable method to indicate responses. The use of letters is also preferable to numbers, since numerical answers in numbered items may be confusing (Gronlund, 1993:59). On completion of the questionnaires, each one was reviewed for omissions. The answers were scored dichotomously (0 or 1).

The questionnaire was administered to five times more respondents than the number of items (37) (as recommended by Huysamen, 1986:46). The students (n=186) who attended the classes allocated to the study were invited to participate. Two groups of undergraduate higher education students participated voluntarily and anonymously. The first group of 99 first- and second-year National Diploma (ND): Consumer Science: Food and Nutrition and ND: Somatology students, who have nutrition as a subject, were expected to have a higher level of nutrition knowledge and formed the knowledgeable group. The second group of 87 first- and second-year ND: Environmental Health and General Education and Training students were expected to have a low level of knowledge in the field of nutrition as these programmes do not include nutrition as a subject. This would ensure that one group had a greater knowledge of nutrition than the other, while other variables such as gender, age and educational level were fairly similar for the groups. The number of students participating also met the number indicated by Nunnally (1972:194) as at least 40 and preferably 100 for efficient item analysis (that is, that there should be at least 40 students tested to obtain the difficulty index and the distribution of percentages for alternatives on multiple-choice items; and 80 students for the discrimination index).

Item analysis

Item analysis involves the statistical analysis of the results of a test administration to identify which items can be retained and which need to be either revised or discarded (Nunnally, 1972:186). The data of the completed questionnaires were entered on a spreadsheet. The item analysis was done using MS Excel and Stata 8. Only those items meeting the analysis criteria were retained for the final test. Gronlund (1993:102) indicated that the item analysis procedure for norm-referenced tests should provide information on item difficulty, discrimination power of item and effectiveness of each alternative.

Item difficulty index The item difficulty index indicates the percentage of respondents who answer an item correctly (Nunnally, 1972:186; Thorndike *et al,* 1991:212). It was calculated so that only items with suitable difficulty indices were included in the test (Nunnally, 1972:187). Although this statistic is frequently referred to as the item difficulty rating, Nunnally (1972:186) sees this as a misnomer because the higher the percentage, the easier – rather than more difficult – the item is (Nunnally, 1972:186). Thorndike *et al* (1991:212) argue that it should be called the "easiness" of an item. For example, an item that 75% of respondents answer correctly would have a difficulty index of 0,75 (Thorndike *et al,* 1991:212).

Although it is better to avoid items that most respondents will answer correctly or incorrectly, it is possible to have an item that many respondents will answer incorrectly because it may assess some useful information, or an easy item that covers a particularly important point that everybody should know (Thorndike *et al,* 1991:245). A good rule is to use few items that are either above 0,80 or below 0,20. For a multiple-choice test consisting of four or more alternatives, items in the range between 0,35 and 0,85 should be selected (Nunnally, 1972:189). This criterion was used in this test. All items found to be too easy (answered correctly by more than 85% of respondents) or proved to be too difficult (answered correctly by less than 35% of respondents) were excluded.

The item difficulty indices were used to rank the items from the easiest to the most difficult (Nunnally, 1972:189) in the final test. Placing easier items at the beginning provide respondents with an optimistic start and placing more difficult items near the end prevent respondents from spending too much time on difficult items early in the testing period (Nunnally, 1972:188).

Distribution of answers to alternatives The percentage of respondents who indicated each of the alternatives (in this case a, b, c or d) as their answer were computed to determine the distracting ability of the alternatives. A useful standard to apply, regardless of the number of alternatives, is to replace those alternatives which were not chosen by at least 5% of the respondents by a more plausible alternative, as that alternative could not be regarded as a good distracter (Nunnally, 1972:190). Items in which alternatives were indicated by 5% or less of the respondents were discarded.

Discrimination index This index is used to determine the extent to which each item measures the same aspect as the total test in which it was included. There are several ways to do this. One of the simplest methods is the discrimination index. It can be described as follows: First find the top and the bottom 25% of the respondents in terms of total test scores (Note: Test results of the middle group are set aside and not used). Next, determine for each item the percentage of students in the top and the bottom groups who answered the item correctly. Finally, subtract the percentage for the bottom group from the percentage for the top group (Nunnally, 1972:192). The discrimination index can also be determined by subtracting the number of students who answered the item correctly in the lower group, from the number of students who answered it correctly in the upper group, and dividing the difference by the respondent number in one group (Thorndike *et al,* 1991:250). The larger the difference, the better the item as it could discriminate the top from the bottom respondents. If the difference is small, the item failed to discriminate between good and bad performers. Although it may be desirable to use the top and bottom 25% if the group is large (or the upper and lower halves if the group is small), selecting the top and bottom 27% is recommended for a more refined analysis (Gronlund, 1993:105). Thorndike *et al* (1991:249) also indicate 27% that was used in this test .

Items for which the difference is not at least 20 percentage points (Nunnally, 1972:192) or which have a discrimination index of below 0,20 are poor discriminators (Thorndike *et al,* 1991:251) and were eliminated.

Item-to-total correlation Each student has a score on each item (e.g. pass or fail) and on the test as a whole. Correlation coefficients can be computed from these results. In addition to the discrimination index, the most popular statistical index to determine the internal consistency of a test is the item-to-total correlation (Nunnally, 1972:193). The Pearson's correlation was used to calculate the item-to-total test correlations. Only items that met the correlation of 0,20 and higher were retained as Parmenter and Wardle (2000) indicate this as the minimum suggested correlation.

Final knowledge test

Those items that met the item analysis criteria formed the final knowledge test. The two most important considerations of a well-constructed knowledge test are validity and reliability (Gronlund, 1993:34).

Validity This is the most important quality to consider and refers to the appropriateness, meaningfulness and usefulness of score interpretations (Gronlund, 1993:159). There are two aspects to validity, namely, what is measured and how consistently it is measured. Consistency refers to the reliability of the scores. Reliability is therefore a necessary ingredient of validity, but it is not sufficient to ensure validity. Validity has to do with the meaning of scores, i.e. do test scores measure what the test user intends them to measure (Ebel & Frisbie, 1991:100) and how are the scores used to make decisions (Ebel & Frisbie, 1991:101). The traditional view that there are several types of validity has been replaced by the view that validity is a single, unitary concept based on various forms of evidence. The former types of validity (content, criterion-related and construct) are now referred to as content-related, criterion-related and construct-related evidence. For some test score interpretations only one or two types of evidence may be critical, but an ideal validation includes evidence from all three categories. In many situations this ideal is not accomplished (Gronlund, 1993:161).

Criterion-related validity refers to the degree to which test scores are related to some other valued measure called a criterion. There are two types of studies used in obtaining criterion-related validity. The first is concerned with predicting future performance (called a predictive study) and the other with estimating current performance (called a concurrent study as both test and criterion are obtained at the same time) (Gronlund, 1993:163). Criterion-related validity is typically expressed by a correlation coefficient as it indicates the degree of relationship between two sets of measures, i.e. the test scores and the criterion (Gronlund, 1993:164). In some cases, appropriate criterion measures are simply not available (Ebel & Frisbie, 1991:107). Criterion validation of nutrition knowledge measurement instruments is difficult because no "gold standard" is available to compare instruments with (Steenhuis *et al,* 1996; Parmenter & Wardle, 2000).

Construct validity (referred to by Axelson & Brinberg (1992) and Sapp & Jensen (1997) as discriminant validity) requires that the construct presumed to be reflected in test scores actually does account for the difference in test performance (Gronlund, 1993:166). This can be achieved by comparing scores of known groups to determine whether the scores differentiate the groups as predicted (Thorndike *et al,* 1991:142; Gronlund, 1993:167). Construct validity was chosen above criterion-related validity as it did not require a criterion. The use of sub-populations with expected differences in nutrition knowledge in order to determine construct validity had been used in the development of nutrition knowledge tests (Steenhuis *et al,* 1996; Parmenter & Wardle, 1999). The Mann-Whitney test was used to determine the construct validity. If the test scores between the student groups were significantly different in the expected direction, it could be said that the test measured what it was suppose to measure, namely a nutrition knowledge dimension, and that it was possible to obtain a certain degree of discrimination between groups that were expected to differ in nutrition knowledge (Steenhuis *et al,* 1996).

Reliability Reliability refers to the consistency of test scores – that is, how consistent they are from one measurement to another (Gronlund, 1993:169) or how error free the measurements are (Gronlund, 1993:176). Reliability is typically reported by means of a reliability coefficient, which is also a correlation coefficient (Ebel & Frisbie, 1991:77; Gronlund, 1993:169). This could be accomplished by administering the same test to a group twice, with a time interval in between (test-retest method), administering two equivalent forms of the test with a time interval in between (equivalent-forms method), or administering the test once and computing the consistency of the responses within the test (internal-consistency method) (Ebel & Frisbie, 1991:81-82; Gronlund, 1993:170). Each one of these methods measures a different type of consistency (e.g. over time, over different samples of items and over different parts of the test) (Gronlund, 1993:176).

The simplest means to estimate internal-consistency of test scores from a single test administration is to use the Kuder-Richardson (K-R) formulas (Gronlund, 1993:171). The K-R formulas are the most widely accepted methods for estimating reliability (Ebel & Frisbie, 1991:83). The K-R formula 21 (K-R21) requires information on the test mean/average score (Gronlund, 1993:171). If items do not vary widely in difficulty, computing the mean is reasonable, but when items vary in difficulty, as they almost always do, the K-R formula 20 (K-R20) that requires information about the difficulty (proportion of correct responses) of each item should be considered. The K-R20 is applicable to tests scored dichotomously (0 or 1) (Ebel & Frisbie, 1991:85). The Cronbach's alpha reliability coefficient may be used in the place of the K-R formulas for estimating the reliability of scores from tests not scored dichotomously (Ebel & Frisbie, 1991:98).

The test-retest method was not chosen as the time-interval between test administrations is an important factor. A short interval can lead to overestimation of reliability, while a longer interval can lead to an underestimation (Steenhuis *et al,* 1996). The equivalent forms method was also not considered as it is difficult to construct two very similar forms of the same test with items the same or equal in difficulty, and respondents may become tired and/or bored during completion. Parmenter and Wardle (2000) argue that this practice is rarely, if ever, applied. Applying the split-half method again does not consider the possibility that there are numerous methods besides on even and uneven numbered items by which the test could be split. The K-R20 considers this possibility (Ebel & Frisbie, 1991:84) and was chosen to determine the reliability or internal consistency of the test.

There are no absolute standards to determine if a reliability coefficient is high enough. Some standards have evolved over time. The generally accepted minimum is 0,65 if scores of a group of individuals are to be determined (Ebel & Frisbie, 1991:87). Fanslow *et al* (1981) consider 0,75 as acceptable and Pirouznia (2001) 0,70.

RESULTS AND DISCUSSION

Demographic data of respondents

The gender and age profile of the respondent groups are indicated in Table 1. The respondents were predominantly female and homogeneous with regard to age within the groups.

Test item analysis

Seventeen of the 37 items (46.0%) were retained after the item analysis. The results of the item analysis are indicated in Table 2. Six items did not meet the difficulty index criterion of 0,35 to 0,85 (or 35 to 85%), of which five items were found to be too difficult (questions 7, 11, 17, 23 and 29) and one item (question 5) too easy. The questions found to be too difficult related to the fibre content of different vegetables (question 7), fruit (apple processed differently, i.e. raw, baked, dried, juice) (question 11) and meat (question 17), as well as the

TABLE 1: DEMOGRAPHIC CHARACTERISTICS OF RESPONDENT GROUPS

Demographic characteristic	Knowledgeable group (n=99)	Less knowledgeable group (n=87)
Gender		
Female	95 (96%)	76 (87%)
Male	4 (4%)	11 (13%)
Age in years		
Median	19 – 20	19 – 20
Average	20	20

TABLE 2: RESULTS OF TEST ITEM ANALYSIS

Item	Difficulty index*	Discrimination index**	Item-to-total correlation***	Distribution of answers to alternatives ****			
				a	b	c	d
1	0,64	0,61	0,524	62,9%	7,5%	6,5%	23,1%
2	0,68	0,48	0,493	15,6%	67,7%	9,1%	7,5%
3	0,64	0,55	0,512	9,7%	62,4%	10,8%	17,2%
4	0,76	0,40	0,559	8,1%	7,5%	77,4%	7,0%
5	**0,88**	0,27	0,501	**1,6%**	9,1%	**2,2%**	87,1%
6	0,39	0,28	0,338	16,7%	30,1%	39,8%	13,4%
7	**0,32**	**-0,02**	**0,107**	29,6%	18,3%	24,2%	28,0%
8	0,58	0,21	0,280	8,6%	20,4%	10,2%	60,8%
9	0,36	**0,18**	**0,142**	34,4%	10,8%	31,7%	23,1%
10	0,59	0,35	0,291	22,6%	6,5%	11,3%	59,7%
11	**0,20**	**0,15**	**0,181**	8,1%	19,4%	8,1%	64.5%
12	0,62	**0,10**	0,230	**3,8%**	31,2%	60,2%	**4,8%**
13	0,85	0,31	0,435	87,6%	**1,1%**	6,5%	**4,8%**
14	0,46	0,43	0,471	18,3%	44,6%	18,3%	18,8%
15	0,74	0,34	0,462	73,7%	18,3%	5,4%	**2,7%**
16	0,56	0,47	0,391	59,1%	20,4%	13,4%	7,0%
17	**0,32**	0,52	0,450	9,1%	29,0%	30,6%	31,2%
18	0,55	0,49	0,467	15,6%	19,4%	10.8%	54,3%
19	0,67	0,49	0,562	64,0%	14,0%	9,1%	12,0%
20	0,84	0,38	0,523	7,0%	**2,7%**	5,9%	84,4%
21	0,59	0,61	0,588	25,3%	10,8%	58,1%	5,9%
22	0,63	0,30	0,457	9,7%	68,8%	20,4%	**1,1%**
23	**0,30**	**0,18**	**0,134**	37,6%	25,8%	16,1%	20,4%
24	0,82	0,34	0,535	82,3%	8,6%	**2,7%**	6,5%
25	0,66	0,32	0,369	7,0%	14,0%	12,4%	66,7%
26	0,76	0,32	0,551	11,3%	76,3%	**3,8%**	8,6%
27	0,40	0,26	0,266	11,3%	12,9%	37,1%	38,7%
28	0,35	0,26	0,320	25,3%	16,7%	22,0%	36,0%
29	**0,28**	0,27	0,238	32,8%	28,5%	18,3%	20,4%
30	0,44	**0,19**	**0,176**	19,9%	34,4%	43,0%	**2,7%**
31	0,80	0,34	0,578	5.4%	81,2%	10,8%	**2,7%**
32	0,77	0,36	0,578	76,3%	12,9%	**3,2%**	7,5%
33	0,66	0,41	0,455	12,9%	6,5%	65,6%	15,1%
34	0,48	0,33	0,312	12,9%	**4,8%**	50,0%	32,3%
35	0,37	0,40	0,389	37,1%	10,2%	36,6%	16,1%
36	0,55	0,47	0,412	53,8%	35,5%	6,5%	**4,3%**
37	0,59	**0,10**	**0,153**	22,0%	11,8%	8,6%	57,5%

Values in bold print: Item does not meet specific item analysis criteria:

* Item difficulty index: 0,35 - 0,85

** Item discrimination index: ≥ 0,20

*** Item-to-total-correlation: ≥ 0,20

**** Distribution of answers to alternatives: ≥ 5%

food source of soluble fibre (question 23) and the recommended servings of fruit to be eaten daily to ensure good health (question 29). Girois *et al* (2001) also report that both the US and Geneva populations were less likely to recognise that red meat was not high in fibre. The question found to be too easy (question 5) related to the choice of bread containing the highest fibre content, of which the response "breads baked with whole wheat flour" was the correct answer. "Breads baked with cake flour" and "breads baked with self-raising flour" were found not to be good response distracters in this question (discussed below).

Seven items (questions 7, 9, 11, 12, 23, 30 and 37) did not meet the discrimination index criterion of equal to and larger than 0,20 (or 20%). Most of these items related to choosing the food item providing the most fibre per portion at the same weight. The items covered different vegetables (question 7) and fruits (question 9), fruit (apple processed differently) (question 11) and breakfast cereals (question 12). Three of these items found to be poor discriminators were also found to be too difficult (questions 7, 11 and 23). Six items (questions 7, 9, 11, 23, 30 and 37) did not meet the item-to-total correlation criterion of equal to or larger than 0,20. All these items also did not meet either the difficulty index or discrimination index criteria or both. Thirteen items did not meet the criterion that a variable (in this case either a, b, c or d) had to be indicated as the correct answer by at least 5% of the respondents; either one variable (n=10 items; questions 15, 20, 22, 24, 26, 30, 31, 32, 34 and 36) or two variables (n=3 items; questions 5, 12 and 13) did not meet this criterion . For example, breads baked with cake flour and breads baked with self-raising flour (question 5 referred to above), and Rice Crispies and cornflakes as breakfast cereal choices to increase dietary fibre intake (question 12) were too infrequently selected as responses. Where foods high in fats and oils were indicated as response options, they were also not selected frequently enough as responses (questions 13, 15, 20, 24, 26 and 34).

The face validity concern in the item review caused the exclusion of items related to fibre terminology, in addition to most items related to fibre functions, characteristics and recommendations. This led to the 37-item test being made up of 75% food-related questions (75,7%; n=28 items) and 25% food- and health-disease-related questions (24,3%; n=9 items). In the final test, only three of these food- and health-disease-related questions were retained (three of the nine items) because they met the item analysis criteria. These items are related to the function of fibre to aid stool elimination (question 33), the potential harmful effect of reduced mineral absorption when consuming too much fibre (question 35) and oat bran as the fibre containing food with the greatest cholesterol-lowering effect (question 27).

The final test consisted of approximately 80% of the items relating to food (82%; n=14 items) and approximately 20% relating to the food and health-disease relationship (18%; n=3 items). Crawford and Baghurst (1990) indicate that there was uncertainty in their Australian population survey about the role of fibre in disease despite widespread media discussion of this topic. Only 42% of men and 53% of women reported that they were aware of any health problems related to the amount of fibre consumed. Twenty-one per cent of men and women respectively indicated bowel cancer and bowel diseases and 15% constipation. Reporting on the 1989 and 1990 CSFII and the companion DHKS, Variyam *et al* (1996) show that the proportion of the sample that had heard about health problems related to how much fibre a person eats was equally split between yes (50,2%) and no (49,8%). In the study by Parmenter *et al* (2000), participants were asked whether they knew of any links between eating more or less of particular foods and major health problems. Approximately two-thirds of the respondents (62,1%) knew of health risks associated with a low fibre intake, with the majority being aware of the specific risk of cancer. Bowel problems were also associated with insufficient fibre by many people, with the most commonly mentioned disorders being bowel cancer and constipation. In the South African study, 38,2% and 28,9% of the urban adult white and black population samples respectively knew that a low fibre intake was related to major health problems, of which stomach problems/constipation were the most frequently mentioned (Peltzer, 2004).

Final knowledge test

The test consisting of the 17 items retained after the item analysis was found to be reliable. The reliability coefficient determined by the K-R20 (internal-consistency) was 0,9147, which met the criterion of 0,75 identified by Fanslow *et al* (1981) as acceptable where groups are compared. The test was also found to be valid as a significant difference ($p < 0,001$) in knowledge was found in the expected direction between the two student groups utilising the Mann-Whitney test ($z = -9,7473$). The median scores of the knowledgeable and less knowledgeable groups were 12 (mean = 12) and 6 (mean = 6,75) respectively.

The test validity thus established could be explained by the fact that the majority of respondents in the knowledgeable group (n=93; 94%) indicated that formal lectures in their academic programme are a major source of nutrition information; most of these students (n=86; 87%) also indicated that they were more knowledgeable about nutrition in comparison to other young adults. In contrast, most of the less knowledgeable group indicated that they are as knowledgeable and/or less knowledgeable about nutrition in comparison to other young adults (n=62; 71%). More than half of this group also identified printed material, such as magazine articles, lay books, advertisements, brochures and public lectures, radio talks and television programmes (n=24; 28%), and family and friends (n=22; 25%) as their major sources of nutrition information.

The remaining 17 items were ranked according to their difficulty indices (see Table 2) from the easiest to the most difficult items in the final test. Question 4 was placed first and question 28 last. A standard score was also calculated for future use as the norm. The norm score could be used to differentiate between respondents at different levels of achievement: as average, a score of 9 (average median score = 9; average mean

score = 9,37); as above average or good, a score of 12 (median score and mean score of knowledgeable group respectively equalled 12) and above; and below average or poor, a score of 6 (median score and mean score of less knowledgeable group equalled 6 and 6,75 respectively) and below. Tests which interpret each student's relative standing among other students or can compare a student's performance with that of others are norm referenced (Gronlund, 1993:12; Thorndike *et al,* 1991:195).

CONCLUSION AND RECOMMENDATIONS

If health promotion is to be successful, it must address the needs of the target audience and take into account their current behaviour, as well as their knowledge and attitudes (Barker *et al,* 1995). Lack of knowledge might be a barrier to healthy eating (Berg *et al,* 2002). To select a healthy diet, consumers must ignore advertisers' coaxing and taste appeal, and draw on complex scientific and technical knowledge concerning nutrients, foods and health. At a minimum, they need to know the nutritional recommendations, be able to apply those to the food products they are considering and combine recommendations to make the best food choices (Parmenter *et al,* 2000). Buttriss (1997) concludes that many consumers do not have the knowledge to translate healthy messages into food choices. Among the public, there is a high awareness of buzzwords (such as fibre), but their knowledge is insufficient to enable them to translate it into healthy food choices (Buttriss, 1997). Health professionals also overestimate the public's knowledge about nutrition issues. For instance, Buttriss (1997) reports that only 22% of professionals consider the public's level of knowledge to be a barrier to changing their eating habits. In addition, there are worrying gaps in health professionals' own knowledge of what the different terms mean. There is also broad scepticism about the media's role in providing healthy eating advice (Buttriss, 1997). Although it might be argued that knowledge plays a limited part in food choice (Parmenter *et al,* 2000), it is an obvious precondition for exercising an informed choice (Barratt, 2001).

There is an evident lack of information about the current level of knowledge about fibre amongst South Africans as the major concern in the test development was related to face validity. In the final test, a minimal number of the food- and health-disease-related questions were retained (only three of the nine items evaluated), which emphasises that the concern may have been warranted. One of these three items retained related to the function of fibre to aid stool elimination (question 33). This item had a difficulty index of 0,66, which indicated that only about two-thirds of the respondent group was aware of this major role of fibre although it is widely used in breakfast cereal advertising. Although the term "constipation" was not used in this item, it was used in another item (question 31) that had a difficulty index of 0,80, indicating that 80% of the respondents could answer the question correctly when the term constipation was used, but that they could not maintain this level of correctness when the description ”aid stool elimination” was used. This may indicate that the role of fibre in the prevention of constipation may not be fully understood.

The knowledge measure developed yielded an instrument that can be useful to others. The final test consisting of the 17 items retained after the item analysis was found to be a valid and reliable measure to determine fibre knowledge related to food items and health-disease associations. The reliability coefficient (0,9147) met the criterion of 0,75 indicated by Fanslow *et al* (1981) as acceptable where groups are compared. The test can therefore be used to compare the fibre knowledge scores of groups, but also for the scores of individuals as it met the reliability coefficient of at least 0,85 informally agreed upon by experts in educational measurement to make decisions about individuals (Ebel & Frisbie, 1991:86). The high reliability may be due to the thorough process of item review and the difficulty indices of most of the retained items being between 0,40 and 0,70 (13 of the 17 items). Extremely easy or extremely difficult items add little information as they only differentiate a few respondents from others (Nunnally, 1972:187).

The major loss of items occurred due to responses not being selected often enough as answers (13 of original 37 items; approximately 35%). In a number of items where foods high in fats and oils were indicated, they were not selected frequently enough as answers (questions 13, 15, 20, 24, 26 and 34). This may suggest that these student respondents understand that foods high in fats and oils are not correct fibre food choices. It is therefore recommended that in the drafting of multiple-choice questions to determine fibre knowledge careful consideration should be given to the responses, and that foods high in fat and oils should not be included.

Studies found women to be more knowledgeable about fibre than men are (Variyam *et al,* 1996; Buttriss 1997; Auld *et al,* 1998; Parmenter *et al,* 2000; Girois *et al,* 2001), generally to be better in middle aged than older or young people (Auld *et al,* 1998; Parmenter *et al,* 2000), and to increase with education (Variyam *et al,* 1996; Auld *et al,* 1998; Parmenter *et al,* 2000; Girois *et al,* 2001), and higher socio-economic status (Variyam *et al*, 1996; Parmenter *et al,* 2000). As this student sample was biased in favour of women who tend to have better nutrition knowledge, it is probable that the test may estimate the level of knowledge of females more favourably than that of males. It may also favour older respondents who tend to have better nutrition knowledge than younger respondents.

REFERENCES

AULD, GW, NITZKE, SA, MCNULTY, J, BOCK, MA, BRUHN, CM, GABEL, K, LAURITZEN, G, LEE, YF, MEDEIROS, D, NEWMAN, R, ORTIZ, M, READ, M, SCHUTZ, H & SHEEHAN, E. 1998. A stage-of change classification system based on actions and beliefs regarding dietary fat and fibre. *American Journal of Health Promotion* 12(3):192-201.

AXELSON, ML & BRINBERG, D. 1992. The measurement and conceptualization of nutrition knowledge. *Journal of Nutrition Education* 24:239-246.

BABBIE, ER. 1975. *The practice of social research*. Belmont, Calif. Wadsworth Publishing Company.

BARKER, ME, THOMPSON, KA & MCCLEAN, SI. 1995. Attitudinal dimensions of food choice and nutrient intake. *British Journal of Nutrition* 74:649-659.

BARRATT, J. 2001. Diet-related knowledge, beliefs and actions of health professionals compared with the general population: an investigation in a community Trust. *Journal of Human Nutrition & Dietetics* 14:25-32.

BERG, MC, JONSSON, I, CONNOR, MT & LISSNER, L. 2002. Relation between fast food choices and knowledge of dietary fat and fiber among Swedish schoolchildren. *Journal of Adolescent Health* 31:199-207.

BUTTRISS, JL. 1997. Food and nutrition: attitudes, beliefs and knowledge in the United Kingdom. *American Journal of Clinical Nutrition* 65(suppl):1985S-1995S.

COLAVITO, EA, GUTHRIE, JF, HERTZLER, AA & WEBB, RE. 1996. Relationship of diet-health attitudes and nutrition knowledge of household meal planners to the fat and fiber intakes of meal planners and preschoolers. *Journal of Nutrition Education* 28(6):321-328.

COTUGNA, N, SUBAR, AF, HEIMENDINGER, J & KAHLE, L. 1992. Nutrition and cancer prevention knowledge, beliefs, attitudes and practices: the 1987 National Health Interview Survey. *Journal of the American Dietetic Association* 92(8):963-968.

CRAWFORD, DA & BAGHURST, KI. 1990. Diet and health: A national survey of beliefs, behaviours and barriers to change in the community. *Australian Journal of Nutrition and Dietetics* 47(4):97-104.

CREMER, SA & KESSLER, LG. 1992. The fat and fibre content of foods: What Americans know. *Journal of Nutrition Education* 24:149-152.

EBEL, RL. & FRISBIE, DA. 1991. *Essentials of educational measurement*. 5th edition. Englewood Cliffs, N.J. Prentice-Hall.

FANSLOW, AM, BRUN, JK & HAUSAFUS, C. 1981. The NATS - Nutrition achievement tests for the elementary grades. *Journal of Nutrition Education* 13:90-92.

GIROIS, SB, KUMANYIKA, SK, MORABIA, A & MAUGER, E. 2001. A comparison of knowledge and attitudes about diet and health among 35- to 75-year-old adults in the United States and Geneva, Switzerland. *American Journal of Public Health* 91(3):418-424.

GRONLUND, NE. 1993. *How to make achievement tests and assessments*. 5th edition. Needham Heights, MA. Allyn and Bacon.

HARNACK, L, BLOCK, G & LANE, S. 1997. Influence of selected environmental and personal factors on dietary behavior for chronic disease prevention: A review of the literature. *Journal of Nutrition Education* 29(6):306-312.

HUYSAMEN, GK. 1986. *Sielkundige meting - 'n inleiding*. Pretoria. Academica.

KRISTAL, AR, BOWEN, DJ, CURRY, SJ, SHATTUCK, AL & HENRY HJ. 1990. Nutrition knowledge, attitudes and perceived norms as correlates of selecting low-fat diets. *Health Education Research* 2:467-477.

LEVY, AS, FEIN, SB & STEPHENSON, M. 1993. Nutrition knowledge levels about dietary fats and cholesterol: 1983 - 1888. *Journal of Nutrition Education* 25:60-66.

NUNNALLY, JC. 1972. *Educational measurement and evaluation*. 2nd edition. New York. McGraw-Hill.

PARMENTER, K & WARDLE, J. 1999. Development of a general nutrition knowledge questionnaire for adults. *European Journal of Clinical Nutrition* 53:298-308.

PARMENTER, K & WARDLE, J. 2000. Evaluation and design of nutrition knowledge measures. *Journal of Nutrition Education* 32:269-277.

PARMENTER, K, WALLER, J & WARDLE, J. 2000. Demographic variation in nutrition knowledge in England. *Health Education Research* 15(2):163-174.

PELTZER, K. 2001. Nutrition knowledge and food choice among black students in South Africa. *Central African Journal of Medicine* 48(1/2):4-8.

PELTZER, K. 2004. Nutrition knowledge among a sample of urban black and white South Africans. *South African Journal of Clinical Nutrition* 17(1):24-31.

PIROUZNIA, M. 2001. The influence of nutrition knowledge on eating behaviour – the role of grade level. *Nutrition & Food Science* 31:62-66.

RESNICOW, K & REINHARDT, J. 1991. What do children know about fat, fibre and cholesterol? A survey of 5 116 primary and secondary school students. *Journal of Nutrition Education* 23:65-71.

SAPP, SG & JENSEN, HH. 1997. Reliability and validity of nutrition knowledge and diet-health awareness tests developed from the 1989-1991 Diet and Health Knowledge Surveys. *Journal of Nutrition Education* 29:63-72.

STEENHUIS, IHM, BRUG, J, VAN ASSEMA, P & IMBOS, TJ. 1996. The validation of a test to measure knowledge about the fat content of food products. *Nutrition and Health* 10:331-339.

TEMPLE, NJ. 1999. Survey of nutrition knowledge of Canadian Physicians. *Journal of the American College of Nutrition* 18(1):26-29.

THORNDIKE, RM, CUNNINGHAM, GK, THORNDIKE, RL & HAGEN, EP. 1991. *Measurement and evaluation in psychology and education*. 5th edition. New York. MacMillan.

VARIYAM, JN, BLAYLOCK, J & SMALLWOOD, DM. 1996. Modelling nutrition knowledge, attitudes, and diet-disease awareness: the case of dietary fibre. *Statistics in Medicine* 15:23-35.

WALKER, ARP, ADAM, F & WALKER, BF. 2001. World epidemic of obesity: The situation in Southern African populations. *Public Health* 115:368-372.