Testing and Measurement in College/University Based Basic First Aid and Emergency Care

Michael D. Ballard, Ed.D., CHES[1]; Robert H. Kirk, H.S.D.[2]; Susan M. Smith, Ed.D.[2]

[1] Department of Public Health, Western Kentucky University; [2] Department of Health and Safety Sciences, University of Tennessee

Corresponding author: Michael D. Ballard, Department of Public Health, Western Kentucky University, 408A Science & Technology Hall, Bowling Green, KY 42101; 502.745.5862 (phone); 502.745.4437 (fax); michael.ballard@wku.edu. **Received June 10; revised December 8, 1998.**

Abstract

A valid and reliable basic first aid and emergency care knowledge test for use at the college/university level was developed. The First Aid and Emergency Care Knowledge Test for College and University Students provides a viable alternative to the existing testing mechanisms and affords the classroom instructor the opportunity to compare the performance of students in his/her respective class to others across the nation. This test also has application as a pretest - posttest to guide classroom instruction.

Introduction

Unintentional injuries are currently the leading cause of death in individuals 1 - 44 years of age (Ventura, Peters, Martin & Maurer, 1997). It has long been assumed that training in first aid and emergency care has the potential to mitigate the results of such unintentional injuries and to raise safety and health awareness. As a result, basic first aid and safety has become an integral segment of the high school and college curricula to provide instruction to populations at high risk.

Poole & Ludwig (1960), conducted an investigation of colleges and universities concerning the offering of first aid courses. The results of the investigation revealed that 83% of the colleges and universities surveyed offered first aid as a separate course, or as part of another course. In addition, Winkelman (1977), stated that "for some time community colleges and four-year colleges and universities have offered first aid and emergency care courses." Consequently, "first aid and emergency care subject matter has also been found to be included in more comprehensive kinds of health and safety education courses such as a general safety and accident prevention course or a course concerning the organization and administration of health programs."

The first aid and emergency care instruction provided at the college and/or university level must be more comprehensive and utilize a more in depth method of evaluation. The majority of the classes provided by the respective training agencies are developed so that minimal effort is required to successfully complete the instructional program. The amount of time required to complete the instructional program varies, but can be obtained in a limited time period. At the collegiate level, basic first aid and emergency care instruction is provided during the course of a semester and/or quarter. The amount of time that is devoted to this instruction is dependent on the respective institution. Despite the variation in the time allotted, instruction at the college level generally provides a more in depth introduction to the topic of first aid and emergency care than is provided in community based training programs.

As with other subject content areas, current instruction is heavily scrutinized with an emphasis placed on accountability. Measurement and evaluation, of which testing is an example, are two tools utilized to address accountability issues and the planning of course content and delivery.

The measuring instruments used with members of the community taking a specific course should not be the same as those utilized at the collegiate level. Testing for the purpose of student certification is a major concern, but is not the only concern present among college instructors. The college instructor must issue a final grade for the course.

According to the literature, first aid and emergency care measurement and evaluation was initially introduced at the senior high school level in 1940 with the development of the **General First Aid Test for Senior High School Students** (McCloy & Young, 1954). Measurement and evaluation at the collegiate level had its inception at Brooklyn College in the early 1940s (Doscher, 1943). As a result of this impetus, cognitive standardized testing in the area of first aid and emergency care at the collegiate level developed sporadically in the ensuing decades (Serdula, 1957; Casperson, 1959; Gilbert & Windsor,

1977; Burckes, 1982). Despite these early efforts, there is a noticeable paucity within the literature pertaining to continued contributions or research since the early 1980s focusing on measurement and evaluation.

Purpose

The purpose of this project was to develop a valid, reliable, and objective first aid and emergency care knowledge test for college and university students.

In order to provide more comprehensive first aid and emergency care instruction at the college/university level, an instrument was needed for consistent evaluation. the outcome of this project consisted of a valid, reliable test with accompanying national norms which provide an instructor with the opportunity to compare the test performance of his/her students to the performance of others across the nation. Such as test could also be used to identify specific areas in the first aid and emergency care program in need of revision or modification.

Procedure

The initial step in the construction of the instrument was the development of the table of specifications. According to Lien (1980), "the table of specifications is a prescriptive guide in regards to the specific content of a test that is being developed." This involved an in depth examination of universally used first aid and emergency care textbooks to develop a list of possible subject content areas that could be included in basic first aid and emergency care instruction at the college and university level. This was accomplished by analyzing the most widely adopted textbooks namely: *First Aid: Responding to Emergencies*, (American Red Cross, 1993), and *First Aid and CPR* (National Safety Council, 1991). A thorough analysis of the two books suggested twelve subject content areas that could be considered for inclusion in college level instruction. These topics included: emergency identification and recognition; basic life support; control of blood loss; shock; cold related injuries; musculoskeletal injuries; poisoning; burns and heat injuries; splinting, dressing, bandaging; specific medical emergencies; first aid kits and supplies; and victim transport.

A national jury panel consisting of forty-four members was requested to assign weights (percent value) to each of the subject content areas provided in relationship to the number of corresponding items on two seventy-five item parallel form instruments. Each member of the panel had prior experience either in the development of a basic first aid and emergency care textbook, program, or in the provision of first aid and emergency care instruction at the college level. Each juror had extensive knowledge and experience from which to make judgements concerning which subject content areas should be included on a basic first aid and emergency care knowledge test. The weightings provided by the panel of jurors were tabulated to ascertain the number of test items that should be devoted to each subject content area. The table of specifications is provided in Table 1.

Table 1. Table of Specifications: Subject Content Area Weightings of National Jury Panel of Experts

Subject Content Area	Percentage Subject Content Should Cover on Test	Number of Test Items per Content Area (75 total)
Emergency Identification & Recognition	11.9	9
Basic Life Support	18.3	13
Control of Blood Loss	9.9	8
Shock	10.0	8
Cold Related Injuries	5.4	4
Musculoskeletal Injuries	9.2	7
Poisoning	5.7	5
Burns and Heat Injuries	6.7	5
Splinting, Dressing, Bandaging	6.6	5
Specific Medical Emergencies	8.6	7
First Aid Kits and Supplies	3.4	2
Victim Transport	4.4	3
Total	100.0	75

Due to the considerable variability in the amount of time scheduled for a class period at the various colleges and universities, it was determined that two equivalent forms containing seventy-five multiple-

choice items each would be appropriate. Most college and university courses range from 45-75 minutes in length. It was concluded that an instrument containing seventy-five items could be completed in those college or university classes that were limited to forty-five minutes in length. This was based on the assumption that the typical lower division college or university student could be reasonably expected to respond to an average of two (2) multiple choice items per minute. Gronlund (1985), stated that "as a rough guide, the average high school student should be able to answer...one multiple choice item...per minute." Using this as a guide, it was assumed by the researchers that due to the fact that the average college student is at an advanced level of intellectual maturity than high school students, that two multiple choice items per minute would be acceptable.

Next, members of the panel of jurors evaluated each test item that was constructed in regards to whether the item should be included, revised, or discarded. Items were then randomly selected from each content area to be placed in the parallel form of the preliminary drafts with respect to the percentages enumerated in the specifications table.

The two preliminary drafts of the instrument were administered as a pilot at twelve institutions to a population of 313 students (n=313). Form A of the preliminary instrument was administered to a student population of 156. Form B was piloted with a sample of 157 students. The 12 participating institutions represented 6 two-year community and/or junior colleges and 6 four-year colleges and/or universities randomly selected from each of the six geographical districts of the American Alliance for Health Physical Education, Recreation, and Dance (AAHPERD).

Results

Concerning validity it must be remembered that "tests themselves are never valid...rather the concept of validity is linked to the inferences we draw based on the use of tests" (Popham, 1990). Rather the professional must make judgements regarding the validity of such score-based inferences. These judgements are based on evidence of which there is three types, of which one was specific to this investigation, content evidence. "In general, content-related evidence demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content" (Popham, 1990). The protocol to ensure this type of validity espoused by Popham (1990) which was followed included: First, attempts to incorporate suitable content on the test can be carried out and (documented) during the test development itself, calling on experts where needed to ensure that the test represents a desired domain of content. Second, is the completion of post facto judgements concerning the representativeness of the content of the test.

The content validity of the instrument and test items were established by the use of two well documented basic first aid and emergency care textbooks in the identification of initial subject content areas. The subject content areas included in the development of the draft instrument were determined through preliminary and final weightings provided by a 44 member panel of experts who possessed either extensive experience or specialization in planning, developing, or teaching basic first aid and emergency care.

The data analysis for pilot instrument Form A revealed the following: the mean was 41.6, the median was 44.1, and the mode was 44; the range was 54 with a variance of 155.05 and a standard deviation of 12.4. The item analysis for pilot instrument Form A revealed that of the seventy-five items, 30 (40%) were at the .41 or higher discrimination level which is considered to be very good. Twenty-seven of the items (36%) were within the .20 to .40 boundaries which were classified as satisfactory. Fifteen items (20%) were within the .00 and .19 level of discrimination. A total of three items were determined to be negative discriminators. The mean level of item discrimination for the total pilot instrument Form A was .34, which was within the satisfactory range of acceptance. The discrimination range utilized was reported by Lien (1980), which classified .41 to 1.00 as very good; .20 to .40 as satisfactory; .00 to .19 as low; and -.01 to -1.00 as unacceptable.

The item analysis revealed that From A contained ten items (13%) that were deemed too easy, with difficulty levels above 75%. The majority of the items, 59 (79%), were within the acceptable level of difficulty which was .25 - 75%. Six items (8%) were too difficult, with a difficulty level below 25%. The difficulty for the total pilot instrument Form A was 55.2% The mean for pilot instrument Form B was 44.1, the median, 46.2; the distribution was bimodal, with the scores of 49 and 47 occurring most frequently.

The range was 47, with a standard deviation of 10.4, and a variance of 109.15.

The item discrimination levels for pilot Form B revealed that thirteen items, or 17%, where at the .41 or higher level of discrimination. Thirty-eight (51%) of the test items were between the satisfactory ranges of .20 and .40. Twenty-one (28%) items were in the low category concerning discrimination. Negative discrimination levels were present for three items constituting 4% of the test items. The mean level of item discrimination for the entire Form B instrument was determined to be .26.

Form B consisted of seventeen items (22%) that were too easy. Fifty-one items, or 68%, possessed acceptable levels of difficulty, whereas seven items, or 9%, were classified as too difficult. The cumulative difficulty for pilot instrument Form B was also calculated to be 60.6%.

Kuder-Richardson Formula 21 was utilized in the calculation of a reliability coefficient. According to Popham (1990), "most well constructed norm referenced tests usually hover between .80 and .90." The reliability of Form A was determined to be .89, and Form B was determined to be .84. Therefore, both instruments possessed a suitable level of reliability to be considered well constructed according to Popham's (1990) stipulation.

The parallel forms of the final instrument, Form A and Form B containing seventy-five items each, related directly to the statistical data provided by the item analysis of the preliminary draft. Only those test items that met the following established criteria were selected to be a part of the final parallel instruments.

1. The items discriminated between good and poor student performance.
2. The difficulty index for each item was between 25% and 75%.
3. Each alternative response was selected by at least three percent of the respondents.
4. Those pilot instrument items that were deficient in distractor utilization or level of difficulty, or discrimination were either revised or deleted.

The final forms of the instrument were administered in order to establish initial national test norms. The sample consisted of two-year community and/or junior colleges, and four-year colleges and/or universities from the six geographical districts of AAHPERD. The final sample of two year institutions represented 17 states and the sample of four year institutions represented twenty-five states. Thirty-two of the fifty states had at least one intact group of students tested as part of the final administration.

The final sample consisted of n = 927 college and university students. As in the administration of the pilot instrument, cluster sampling was employed, whereby the parallel forms of the instrument were administered to an intact class of students enrolled in a basic first aid and emergency care course at the respective institution.

The analysis of data followed the same procedure as for the pilot forms of the instrument. Percentile ranks. Analysis of the data showed that the final forms of the instrument met acceptable criteria concerning test construction protocols. The measures of central tendency for final Form A were: mean 43.1, median 43.5, and mode 41. The measures of variability for Form A included a range of 52, standard deviation of 9.7, and a variance of 94.09. The Kuder Richardson reliability coefficient was determined to be .82, with an index of discrimination of .33, and an index of difficulty of 57%.

The measures of central tendency for final Form B included a mean of 41.6, median of 41.7, and a mode of 38. The measures of variability included a range of 49, standard deviation of 9.8, and a variance of 95.05. The Kuder Richardson Reliability coefficient was calculated to be .82, the index of discrimination was .34, and the index of difficulty was 52%. The standard error or measurement for Form A was 4.1, and 4.3 for Form B.

The reliability coefficient of the final instruments was calculated using the Kuder-Richardson Formula 21. The reliability for both Final Form A and Form B were calculated to be the same at .82, thus meeting the criteria between .80 and .90 for a well developed, norm-referenced test. The standard error of measurement was 4.1 in Form A, and 4.3 in Form B. In addition, t-scores were calculated for the parallel forms of the final instrument.

Discussion

The First Aid and Emergency Care Knowledge Test for College and University Students (Ballard, 1994), lends itself to utilization as a pretest - posttest instrument to be used in the planning of course content delivery. Such utilization would assist in the identification of areas in which the student(s) is strong/weak to guide instruction. It is believed that if instruction addressed the subject content areas in

which the student(s) had misconceptions and/or the greatest need, the student would potentially exhibit a higher level of interest and motivation in learning and feel comfortable in applying the knowledge in a real world emergency or situation. This is important in response to a 1998 Annals of Emergency Medicine article which revealed that almost 50% of adults surveyed would not provide assistance to someone involved in a roadside accident (Braslow & Brennan, 1998).

This instrument provides a viable alternative to the existing testing mechanisms and affords the classroom instructor the opportunity to compare the performance of students in his/her respective class to a national norm. The instrument is applicable at both the two-year and four-year college and university setting and encompasses the guiding principles of the leading first aid and safety organizations. The instrument is also one which can be administered in a variety of university class settings. As a result of the instrument having 75 multiple choice items it is comprehensive in scope, yet can be administered in classes that range from 45 to 75 minutes. We as a profession must continuously strive to develop measurement and evaluation tools applicable in the college and community setting. Only through such diligent efforts can we assure that the instruments utilized are valid and reliable for the populations in question. In addition, numerous changes in first aid and emergency care theory and practice occur, and must be incorporated in such educational testing devices.

References

American National Red Cross. (1993). *First aid: Responding to emergencies*. St. Louis: Mosby-Year Book, Inc.

Ballard, M. D. (1994). Development of a valid and reliable first aid and emergency care test for college and university students. Unpublished doctoral dissertation, University of Tennessee, Knoxville.

Braslow, A. & Brennan, R. T. (1998). Layperson cpr. *Annals of Emergency Medicine*, 31(5), 654.

Burckes, M. E. (1982). The Burckes Emergency Care Knowledge Test. *Journal of the American College Health Association, 30*(4), 185-186.

Casperson, D. G. (1959). A first aid and emergency care knowledge test for college students (Doctoral dissertation, Indiana University, 1959). *Dissertation Abstracts International, 31*, 7379B.

Doscher, N. (1943). Two first-aid examinations for college students and adult groups. *Research Quarterly, 14*(2), 228-248.

Gilbert, G. G., & Windsor, R. (1977). The Ohio State University standard first aid and personal safety achievement test. *Journal of School Health, 47*(1), 48-50.

Gronlund, N. E. (1985). *Measurement and evaluation in teaching* (5th ed.). New York: Macmillan Publishing Company.

Lien, A. J. (1980). *Measurement and evaluation of learning* (4th ed.). Dubuque, Iowa: W.C. Brown Co.

McCloy, C. H., & Young, N. D. (1954). *Tests and measurement in health and physical education* (3rd ed.). New York: Appleton-Century-Crofts, Inc.

National Safety Council. (1991). *First aid and CPR*.. Boston: Jones and Bartlett Publishers.

Poole, R. D., & Ludwig, D. J. (1960). First aid instruction in colleges and universities. *Journal of School Health, 30*, 237-239.

Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective*. Englewood Cliffs, New Jersey: Prentice Hall.

Serdula, G. (1957). *Standardization of a first aid knowledge test for college students*. Unpublished doctoral dissertation, Indiana University, Bloomington.

Ventura, S. J., Peters, K. D., Martin, J. A., & Maurer, J. D. (1997). Births and deaths: United States, 1996. *Monthly vital statistics report*, vol 46 no 1, supp 2. Hyattsville, MD: National Center for Health Statistics. 1997

Winkelman, J. L. (1977). An-up-to-date view of cardiopulmonary resuscitation instruction in colleges and universities. *Journal of the American College Health Association, 26*(2), 108-110.