

Development and Evaluation of a Questionnaire to Assess Physical Educators' Knowledge of Student Assessment

*Kyriaki Emmanouilidou, Vassiliki Derri,
Nicolaos Aggelousis, Olga Vassiliadou*

Abstract

The purpose of this pilot study was to develop and evaluate an instrument for measuring Greek elementary physical educators' knowledge of student assessment. A multiple-choice questionnaire comprised of items about concepts, methods, tools, and types of student assessment in physical education was designed and tested. The initial 35-item questionnaire was revised for its content and face validity, and 22 items remained. Then, 98 elementary physical educators completed the latter version of the questionnaire within half an hour and in the presence of the researcher. After item analysis and reliability test, adequate difficulty and discrimination indexes, acceptable internal consistency and strong test-retest stability were demonstrated for the final 17-item questionnaire. It is concluded that Greek elementary physical educators' knowledge of student assessment can be evaluated with a brief, valid, and reliable instrument.

To educate students effectively is to provide them with the opportunity to reach their potential and contribute to society. This can only happen if teachers and others who work with students obtain data and information about how each student is progressing, what the student knows and can do, and what must be done to further

Drs. Kyriaki Emmanouilidou, Vassiliki Derri, Nicolaos Aggelousis, and Olga Vassiliadou are affiliated with Democritus University of Thrace, Greece.

develop and encourage the student's educational development. In other words, teachers must assess students. According to National Association for Sport & Physical Education (NASPE) (2004), student assessment is the process of gathering evidence about children's level of achievement in a specific area and of making inferences based on that evidence for a variety of purposes. Moreover, Alberta Education (2006) supports that student assessment helps all who are engaged in the educational process make more informed decisions. Student assessment is one of the standards that NASPE (2009) produced to describe what knowledge and skills are essential for effective teaching in physical education from the beginning of teaching career. According to the particular standard, physical educators have to understand and use the varied types of assessment and be aware of their contribution to overall program continuity and the development of the physical, cognitive, social and emotional domains. Also, in the *Physical Educator Book* developed by the Greek Ministry of National Education (2006) student assessment is considered an integral part of the learning process, which can be used for learning evaluation, student motivation and grading, and program adaptation.

In other countries, in teacher exams like the Praxis exams in USA, there are specific topics that should be covered by the physical educator's knowledge on student assessment. In Greece such criteria do not exist. Additionally, in the exams of the Supreme Council for Civil Personnel Selection (ASEP) for teachers' appointment in Elementary and Secondary Education, only two or three of the items refer to student assessment issues and these are not specified for physical education but apply to all science teaching areas.

In Greece, even though a great number of knowledge questionnaires have been developed in the areas of medicine, nutrition and student knowledge in various subjects, there is no such instrument that can be used to assess physical educators' knowledge of student assessment. As Mood (1989) mentions, the purpose of construction and administration of a knowledge test, is to compare an individual's performance to a particular content area. Furthermore, understanding the knowledge basis of teachers is an important task in working toward the construction of adequate content for further teacher education, development and operation in class. Taking into account the importance of student assessment, a valid and reliable test which measures teachers' knowledge in that area should be a useful instrument for the teachers themselves, the

designers of professional development programs and researchers as well. So, the purpose of the present study was to develop and evaluate a test for measuring Greek elementary physical educators' knowledge of student assessment.

Method

Questionnaire Development

A multiple-choice questionnaire was designed to measure physical educators' knowledge of student assessment. The format of multiple-choice questions was selected for many reasons. First of all, according to Haladyna and Downing (1989), multiple-choice items remain the most widely and commonly used item format because, in comparison to other item formats, it has a lower cost and greater efficiency in using and storing items, and is simply too compelling to disregard. In addition, more multiple-choice items can be administered in a given time frame than any other item format. Moreover, it is easy to grade and discourages guessing more than other formats. As a consequence, the reliability of the test data can be increased, a better content sampling can be obtained, and validity can be improved (Hastand & Lacy, 1998). Standards, guidelines, rules and examples of multiple choice questions suggested by researchers (Haladyna, Downing & Rodriguez, 2002; Hastad & Lacy, 1998; Kirkendall, 1987; Mood, 1989; Safrit & Wood, 1995) were followed in order to write adequate stems and distractors.

Each test item of the present questionnaire contains a stem and five alternative answers (a-e), one correct, three incorrect (distractors) with sentence-completion format and an "I do not know" option. The stem was an incomplete sentence, no longer than one and a half line. Researches suggest the "I do not know" option because it reduces guessing and increases internal consistency (Pennington, Pachana & Coyle, 2001).

The four alternatives (correct and distractors) were approximately of the same length and conformed satisfactorily to the stem. Also, there were not alternatives like "none of the above" or "all of the above" because researchers recommend against the use of them (Dudycha & Carpenter, 1973; Haladyna et al., 2002; Harasym, Leong, Violato, Brant, & Lorscheider, 1998) due to the complexity of their effects on the items' ability to discriminate participants and the subsequent reduction of test reliability.

Participants

For the internal validity and reliability examination of the questionnaire, 98 elementary physical educators (50 males and 48 females) with experience of 1 to 24 years ($M=9.6$) from all over Greece participated. They were randomly selected to complete the questionnaire within half an hour in the presence of a researcher during two conferences, one Pan-Hellenic in Thessaloniki and one International in Komotini.

Procedure

The development of the questionnaire proceeded through the following stages: After reviewing the literature, examining course outlines, and analyzing textbooks, a pool of 45 items were developed by the researchers, concerning knowledge, comprehension, analysis and application of student assessment. Initially, the items were true-false and open answered. Thirty-five of them finally became multiple-choice items. The objectives of the questions were general knowledge, elements, types, techniques and tools of student assessment in physical, cognitive, and social domains.

Face and content validity. For content and face validity, two pilot studies were conducted. In the first, a panel of five experts reviewed the 35-item questionnaire and provided comments on the content, language, clarity, importance and appropriateness of items. The panel consisted of three University Professors and two Ph.D. holder physical educators. In the second pilot study 15 in-service physical education teachers, other than the study's participants, who had been randomly selected from three different Greek regions, participated. They provided comments on the comprehension of items and alternatives via individual discussions.

The next stage was to obtain results from a large sample of elementary physical educators so that each item could be tested. Each item of the questionnaire yields one point for a correct answer and a zero for an incorrect or an "I do not know" answer.

Item analysis. The internal validity of the questionnaire was assessed by conducting an item analysis in order to determine the usefulness of each item as a part of the total test (the effectiveness of the examination). Three aspects of item analysis were calculated: a) the item difficulty index which informs about the percentage of participants who correctly answered any given item and a range of .30 to .70 is recommended (Safrit & Wood, 1998) b) the item discrimination index, namely, the degree to which an item separates

the students who did well from those who did poorly. The discrimination index of each item measures the difference between the percentage of participants in, the high third of the sample who correctly answered that test item and the percentage in the low third who correctly answered the item. For each item, values of $>.20$ are considered quite good discriminators (Kirkendall, 1987; Safrit & Wood, 1998) c) the functioning of all possible responses to a test question using “the 2% rule” (Haladyna, 1999; Compton, Quintero, & Esterberg, 2007) to identify distractors that were selected by $<2\%$ of the sample. These kinds of distractors, which have to be eliminated, are likely to be so implausible to participants that hardly anyone selects them.

Reliability. The reliability of the questionnaire was calculated with two indicators: internal consistency and test-retest reliability. Internal consistency was accomplished using Cronbach’s alpha coefficient and the point biserial correlation (pbsc) which concerned the correlation of each item with the whole test. Alpha values of as low as $.50$ are satisfactory for short tests (10-15 items) though tests with over 50 items should yield Cronbach’s a values higher than $.80$ (Kehoe, 1995). Concerning pbsc, values of more than $.15$ mean that the item is behaving as expected, with recommended values between $.30$ - $.70$ (Field, 2005; Kehoe, 1995). The test-retest reliability was assessed by administering the questionnaire to 26 of 98 participants who repeated the questionnaire after an interval of two to three weeks. During this interval the physical educators did not receive any further training on student assessment issues. The test-retest reliability of the questionnaire was assessed using the kappa measure of agreement for each item and the intra-class correlation coefficient for the entire questionnaire.

Results

Content and Face Validity

According to the recommendations of the experts’ panel, 13 items were eliminated and some of the others were modified. Then, 15 private interviews with 15 elementary physical educators were held, focusing on confusion and poor comprehension of stems and alternatives. The procedure resulted in a number of modifications in terms of phrasing in four of the alternatives.

Item analysis

The latter questionnaire contained 22 items and was completed by 98 elementary physical educators. Data was analyzed with SPSS 14 for Windows. The item difficulty of the questions ranged between .24 and .72 except for two items with indexes .12 (too difficult) and .88 (too easy) respectively, which were eliminated. The item discrimination of the questions ranged from .23 to .60 except for three items with .02, .013 and .16 indexes respectively which were removed. Indexes of difficulty and discrimination of the items are showed in Table 1. Additionally, individual distractors were examined using “the 2% rule” and all distractors were acceptable.

Table 1

Indexes of Item Analysis and Internal Consistency of the Questionnaire

Item No.	Index of of difficulty	Index of discrimination	Corrected Item-Total Correlation
1	.53	.37	.288
2	.60	.37	.260
3	.35	.38	.239
4	.56	.23	.203
5	.88	removed	-
6	.41	.35	.230
7	.12	removed	-
8	.56	.34	.182
9	.64	.02	removed
10	.41	.29	.128
11	.24	.33	.348
12	.44	.46	.267
13	.29	.44	.334
14	.28	.30	.256
15	.65	.39	.242
16	.72	.16	removed
17	.68	.42	.242
18	.62	.13	removed
19	.43	.29	.248
20	.63	.30	.228
21	.35	.49	.415
22	.54	.60	.396

Reliability

The internal consistency of the questionnaire was quantified using the responses of 98 participants on the 17-items questionnaire. An acceptable value of Cronbach's alpha was demonstrated ($\alpha=.67$) and the corrected point biserial correlations were above .15 for all items (Table 1). Regarding test-retest reliability, as a method of checking the stability of the test, a strong intra-class correlation was observed ($R=.911$, $p<.05$) with 95% confidence interval from .89 to .98. The measure of agreement (k), using the number of correct and wrong answers of participants in the test and two to three weeks later retest in each item showed a statistically significant agreement ($p<.05$) between tests in all items (Table 2).

Thus, the final questionnaire included 17 items related to techniques (6), tools (2), concepts (3), elements (2) and types (5) of student assessment. The 17-item questionnaire is presented in the Appendix.

Table 2

Number of Correct and Wrong Answers and Measure of Agreement (k) between Tests

Items No.	TEST		RETEST		AGREEMENT
	wrong	correct	wrong	correct	k
1	11	15	13	13	.846**
2	11	15	13	13	.856**
3	13	13	13	13	.692**
4	12	14	11	15	.766**
6	12	14	12	14	1.00**
8	8	18	8	18	.639**
10	13	13	14	12	.769**
11	18	8	18	8	.819**
12	14	12	15	11	.766**
13	17	9	15	11	.839**
14	17	9	17	9	.830**
15	5	21	8	18	.496*
17	8	18	11	15	.428*
19	9	17	12	14	.606**
20	4	22	5	21	.598*
21	11	15	14	12	.620**
22	8	18	8	18	.639**

Note: ** $p\leq.001$ * $p<.05$

Discussion

The enhancement of teachers' knowledge in student assessment positively influences teaching effectiveness in physical education. Even though the importance of the assessment process is generally acceptable, it has received little research attention. The present study aimed to produce a brief questionnaire for Greek elementary physical educators, which evaluates their knowledge of student assessment. The findings showed that the final 17-item questionnaire demonstrated good validity and stability and acceptable internal consistency.

When the questionnaire was designed, several knowledge testing methods were considered. The multiple-choice method was chosen because it is easy to grade and not time-consuming. The "I do not know" option was included in order to avoid guessing. To ensure the content validity the pool of questions/items was constructed and checked by highly experienced professors and teachers specialising in the area of physical education. In all the phases of the questionnaire development there was a careful wording of the questions and specific instructions for completing the questionnaire, where participants were informed of the anonymity and confidentiality of data.

Concerning internal validity, the value of each item in contributing to the purpose of the questionnaire was determined by three criteria: item difficulty and item discrimination indexes and item function. The final format of the knowledge test had 17 items that fulfilled the above criteria according to the Kirkendall (1987) and Safrit and Wood (1995) suggestions.

Regarding the internal reliability of the questionnaire, each item should correlate more than .15 with the total test score, otherwise it does not measure the same ability as does the whole test (Kehoe, 1995). In the present study, the above correlation exceeded the suggested index for all 17 items. Additionally, Cronbach's alpha coefficient, which provides a measurement of internal consistency of the test items, was .67 which is regarded as satisfactory (Field, 2005; Kehoe, 1995). In regard to questionnaire stability over time, it was examined with the test-retest method within a period of few weeks. Good test-retest reliability was demonstrated with high intra-class correlation and significant agreement between initial and repeated scores.

Instruments like this may be important to evaluate what teachers know, because wrong answers and use of the “I do not know” option indicate a knowledge deficit revealing the need of training in the particular subject area. On the other hand, student assessment is a large domain of knowledge and understanding. In the present study a first step was attempted in developing a valid and reliable instrument which evaluates knowledge of student assessment. The short number of items should be recognized as a limitation of the study and further studies should be done in order to extend it.

It is worth mentioning that most of the participants expressed interest in the area and a willingness to improve their knowledge of it. The present study has formed the basis to identify problem areas so that physical education teachers can be trained on student assessment. Even though the present study did not investigate questionnaire sensitivity in knowledge development after a training programme, future research to this direction is planned. Also, further areas of study could be including questionnaire validation in a different population such as secondary physical educators and the association between physical educators’ assessment knowledge and their practice in class.

To conclude, this study provides support for the use of a brief valid and reliable instrument in Greece to assess knowledge of physical education teachers in methods, types, concepts and tools for student assessment. The data analysis has shown that it consists of items with satisfactory indexes of difficulty and discrimination, internal consistency and stability.

References

- Alberta Education. (2006). *Effective student assessment and evaluation in the classroom*. KNOWLEDGE + SKILLS + ATTRIBUTES. Edmonton, AB: Author.
- Compton, M. T., Quintero, L., & Esterberg, M. L. (2007). Assessing knowledge of schizophrenia: Development and psychometric properties of a brief, multiple-choice knowledge test for use across various samples. *Psychiatry Research*, *151*, 87-95.
- Dudycha, A. L., & Carpenter, J. B. (1973). Effects of item formats on item discrimination and difficulty. *Journal of Applied Psychology*, *58*, 116-121.

- Field, A. (2005). *Discovering statistics using SPSS*. (2nd ed.). London: Sage.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 37-50.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Harasym, P. H., Leong, E. J., Violato, C., Brant, R., & Lorscheider, F. L. (1998). Cuing effect of “all of the above” on the reliability and validity of multiple-choice test items. *Evaluation and the Health Professions*, 21, 120-133.
- Hastand, D. N., & Lacy, A. C. (1998). *Measurement and evaluation in physical education and exercise science*. (3rd ed). Boston: Allyn & Bacon.
- Kehoe, J. (1995). Basic Item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation*, 4(10). Retrieved October 19, 2008 from <http://PAREonline.net/getvn.asp?v=4&n=10>
- Kirkendall, D. R. (1987). *Measurement and evaluation for physical educators*. Champaign, IL: Human Kinetics.
- Ministry of National Education and Religious Affairs Pedagogic Institute (2006). *Physical education, 5th and 6th grades. Educator’s book*. Athens: O.E.D.B. (Published in Greek language).
- Mood, D. P. (1989). Measurement methodology for knowledge tests. In M. J. Safrit & T. M. Wood (Eds.), *Measurement concepts in physical education and exercise science* (pp. 251-269). Champaign, IL: Human Kinetics.
- National Association for Sport and Physical Education. (2004). *Moving into the future: National standards for physical education* (2nd ed.). Reston, VA: Author.
- National Association for Sport and Physical Education (2009). *National standards and guidelines for physical education teacher education* (3rd ed.). Reston, VA: Author.
- Pennington, H. R., Pachana, N. A., & Coyle, S. L. (2001). Use of the facts on aging quiz in New Zealand: Validation of questions, performance of a student sample, and effects of a don’t know option. *Educational Gerontology*, 27(5), 409 – 416.

Safrit, M. J., & Wood, T. M. (1995). *Introduction to measurement in physical education and exercise science*. (3rd ed). USA: Mc Graw-Hill.

Appendix

Teachers' Questionnaire of Student Assessment

This questionnaire has been designed to assess the level of knowledge you have about student assessment. You are not expected to know the answers to all the questions. If you do not know the answer or are unsure about it, please mark the *I do not know* option. The information obtained will be treated in the strictest confidence and used only for research purposes. Please circle the one answer you think is correct. THERE IS ONLY ONE CORRECT ANSWER.

The term "student assessment" means ...

- a) collecting data for grading
- b) correcting attitudes and behaviours
- c) collecting and interpreting information
- d) correcting errors in performance
- e) I do not know

When students are evaluated with a norm-referenced system, their individual scores are compared with...

- a) the correct performance of the skill
- b) their previously obtained scores in that test
- c) the average score of a group of comparable students
- d) the best national score
- e) I do not know

Cognitive elements in which students should be evaluated are...

- a) motor skill cues, training principles, cooperation
- b) games strategies, sports regulations, cooperation
- c) game strategies, nutrition principles, skill performance
- d) game rules, cooperation principles, fitness concepts
- e) I do not know

For the quantitative assessment of 3rd grade students in dribbling you...

- a) set the time within which they must complete a performance in dribbling slalom
- b) ask half the students to perform and the others to record the correct cues
- c) dribble and ask students to correct your possible mistakes
- d) ask questions about the cues of skill performance
- e) I do not know

Tools which are used when alternative assessment is applied are ...

- a) checklists and rubrics
- b) standardized fitness and skill tests
- c) multiple-choice tests of sports' history and regulations
- d) traditional games and sequence of skills
- e) I do not know

Domains in which students should be assessed apart from psychomotor are...

- a) cognitive and affective
- b) music-movement and affective
- c) cognitive and interdisciplinary
- d) interdisciplinary and music-movement
- e) I do not know

Teachers make criteria-referenced judgements in order to...

- a) determine students' performance against an established standard
- b) stream students in their class
- c) compare students' performance with their previous ones
- d) assess students' effort
- e) I do not know

A technique of authentic student assessment is ...

- a) observation of performance during drills
- b) standardized test of sport skills

- c) written test of understanding sport skills and strategies
- d) observation of performance in real-life situations
- e) I do not know

Motor elements in which students should be evaluated are...

- a) fundamental skills, strategies, regulations
- b) fundamental skills, fitness, dance
- c) strategies, cooperation, obedience to rules
- d) fitness, sports skills, regulations
- e) I do not know

The formative assessment monitors students' performance ...

- a) during a module
- b) at the end of the term
- c) at the end of a module
- d) at the beginning of a module
- e) I do not know

Techniques of informal student assessment are...

- a) student interviews, oral questions, logs
- b) checklists, regulations tests, sport skills tests
- c) record of progress, written questions, logs
- d) oral questions, written questions, cognitive tests
- e) I do not know

The main purpose of student assessment is ...

- a) detection of student level compared to others
- b) checking of class progress
- c) detection of student level and grading
- d) checking of student and teacher effectiveness
- e) I do not know

The assessment of students' attitudes during the course is made by completing ...

- a) a motor skills' checklist by the teacher
- b) a scorecard by students
- c) a values rating scale by the teacher
- d) a log by students
- e) I do not know

To assess the students' knowledge the teacher uses techniques such as...

- a) matching tests, oral questions, role playing
- b) attitudes surveys, fitness tests, motor skill tests
- c) scorecard completion, oral questions, fitness tests
- d) multiple choice tests, written questions, problem solving
- e) I do not know

Cooperation during a course can be assessed by...

- a) the motor performance of students in each group
- b) the design of a game from each student group
- c) student performance in a traditional game
- d) the praise of every effort by the whole class
- e) I do not know

For the cognitive assessment of fifth grade students at the end of a basketball unit you ask them to...

- a) play 3X3 and you record the score of the game
- b) perform skill drills and you record their performance
- c) play 3X3 and you record their skills performance
- d) answer questions about basketball regulations
- e) I do not know

It is not benefit of the peer-evaluation method the....

- a) development of critical thinking
- b) enhancing of cues' learning
- c) improvement of cooperation
- d) facilitation of teacher
- e) I do not know