IJASRM

International Journal of Advanced Scientific Research and Management, Volume 4 Issue 5, May 2019

www.ijasrm.com

ISSN 2455-6378

# Developing and Validating a Questionnaire in the Field of Physical Education and Sports Sciences

**Poorva Rastogi[1] And Kalpana Sharma [2]**

[1]Research Scholar, Amity School of Physical Education & Sports Sciences, Amity University, Noida, India

[2]Supervisor & Director, Amity School of Physical Education & Sports Sciences, Amity University, Noida, India

**Abstract**

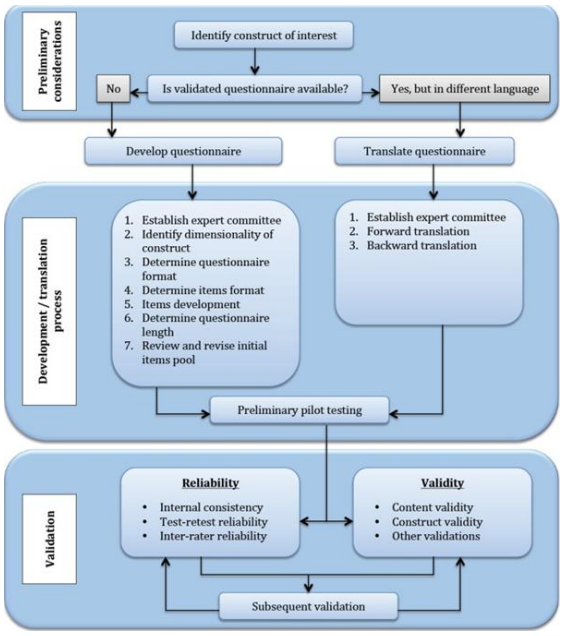
The task of developing a new questionnaire might be overwhelming. The greatest challenge perhaps is to come up with a questionnaire that is statistically sound and is efficient and effective for use in research. This paper provides guidelines and the process for the development of questionnaires for the assessment of effectiveness of physical education curriculum for school children. I would like to provide a framework to guide researchers through the various stages of questionnaire development and validation. To ensure that the questionnaires are psychometrically sound. Several statistical methods are presented to assess the reliability and validity of the questionnaires.

**Keywords:** *Sport sciences, development, questionnaires, validation*

## 1. Introduction

Questionnaires or surveys are widely used in the field of Physical Education and Sports Sciences researches to collect quantitative information from both the student and teaching professionals. Data of interest could range from observable information (e.g., active level, eating habits etc.) to the athlete' subjective training of their status (e.g. physical strength, endurance psychological status). Although using an existing questionnaire will save time and resources, a questionnaire that measures the construct of interest may not be readily available. As a result, investigators may need to develop a new questionnaire as per their respondents. This paper aims to provide straightforward measure for the development of questionnaires (or scales) for the assessment of effectiveness of physical education curriculum for school children in achieving the physical education objective for readers who may be unfamiliar with the process of questionnaire development and validation.

This paper is divided into two main sections. The first section discusses issues that investigators should be aware of in developing a questionnaire. The second section of this paper illustrates procedures to validate the questionnaire after the questionnaire is developed. A model for the questionnaire development and validation process is presented in Figure 1.




## 2. Preliminary Considerations

It is crucial to identify the construct that is to be assessed with the questionnaire, as the domain of interest will determine what the questionnaire will measure. The next question is: How will the construct be operationalized? In other words, what types of behaviour will be indicative of the domain of interest? Several approaches have been suggested to help with this process, such as content

analysis, review of research, critical incidents, direct & indirect observations, expert advice, and instruction.Once the construct of interest has been determined, it is important to conduct a literature review to identify if a previously validated questionnaire exists. A validated questionnaire refers to a questionnaire/scale that has been developed to be administered among the intended respondents. The validation processes should have been completed using a representative sample, demonstrating adequate reliability and validity. If no existing questionnaires are available, or none that are determined to be appropriate, it is appropriate to construct a new questionnaire.

## 3.Developing a Questionnaire

To construct a new questionnaire, several issues should be considered even before writing the questionnaire items.

### Identify the dimensionality of the construct

Many constructs are multidimensional, meaning that they are composed of several related components. To fully assess the construct, one may consider developing subscales to assess the different components of the construct. Next, are all the dimensions equally important? or are some more important than others? If the dimensions are equally important, one can assign the same weight to the questions (e.g., by summing or taking the average of all the items). If some dimensions are more important than others, it may not be reasonable to assign the same weight to the questions. Rather, one may consider examining the results from each dimension separately.

### Determine the format in which the questionnaire will be administered

Will the questionnaire be self-administered or administered by a research/PE professional? This decision depends, in part, on what the questionnaire intends to measure. If respondents are to complete the questionnaire by themselves, the items need to be written in a way that can be easily understood by most of the respondents (e.g. PE curriculum address the needs of the child? (physical, mental, social, intellectual and emotional).

### Determine the item format

Will the items be open ended or close ended? Questions that are open ended allow respondents to elaborate upon their responses. As more detailed information may be obtained using open-ended questions, these items are best suited for situations in which investigators wish to gather more information about a specific domain. However, these responses are often more difficult to encryption and score, which increases the difficulty of interpretation of individuals' responses.

Questions that are close ended provide respondents a limited number of response options. Compared to open-ended questions, these items are easier to administer and analyse. In this study close - ended items was used with Likert-type scales. Scale anchors are to be used to indicate the degree of agreement (e.g., strongly agree, agree, neutral, disagree, strongly degree). To make use of participants' responses for subsequent statistical analyses, researchers should keep in mind that items should be scaled to generate sufficient variance among the intended respondents.

### Item development

Several guidelines have been suggested for writing items. Items should be simple, short, and written in language familiar to the target population. After reviewing 33 items written in simple and short statement for the questionnaire development. These items assess only a single issue. Items that address more than one issue, or "double-barrelled" items (e.g., "My eating pattern and sleep cycle are affected by daily physical exercise routine."), was not used. Avoid leading questions as they may result in biased responses.

### Determine the intended length of questionnaire

There is no thumb rule for the number of items that make up a questionnaire. The survey ought to contain enough things to quantify the build of interest, however not so long, to the point that respondents experience exhaustion or loss of motivation in completing the questionnaire. A questionnaire should not only possess the most parsimonious (i.e., simplest) structure, but it also should consist of items that adequately represent the construct of interest to minimize measurement error. Although a simple structure of questionnaire is recommended, a large pool of items is needed in the early stages of the questionnaire's development as many of these items might be discarded throughout the development process.

### Review and revise initial pool of items

After the initial pool of questionnaire items are written, qualified experts should review the items. And after the expert review 5 variables was shortlisted i.e. Relevance, Efficiency, Effectiveness, Sustainability & Impact with 33 items. Specifically, the items should be reviewed to make sure they are accurate, free of item construction problems, and grammatically correct.

The selected variables based on review of literature and expert opinion for the study was:

**Relevance**: Refers to the extent to which the objectives of the programme correspond with the requirements and needs of the child. Also refers to the usefulness of a programme.
**Efficiency:** A measure of the relationship between the outputs and the inputs of a programme. Measures how economically resources have been converted into results.
**Effectiveness:** Refers to the extent to which the programme objectives have been achieved. Also refers to when the outputs produce the desired outcomes.
**Sustainability:** Refers to the persistence of a programme. It addresses the long- term effects of a programme and it is concerned with the maintenance, financial and economic viability of keeping a programme running.
**Impact:** Refers to what has happened because of a programme. Looks at the difference the programme has made to the beneficiaries.

## 4.Preliminary pilot testing

Before conducting a pilot testing of the questionnaire on the intended population, it is advisable to test the questionnaire items on a small sample (about 30–50) of respondents. This is an opportunity for the questionnaire developer to know if there is confusion about any items, and whether respondents have suggestions for possible improvements of the items. Researcher can also get a rough idea of the response distribution to each item, which can be informative in determining whether there is enough variation in the response to justify going forward with a large-scale pilot test. This questionnaire was tested on 60 respondents. The questionnaire items should be revised upon reviewing the results of the preliminary pilot testing. This process may be repeated a few times before finalizing the final draft of the questionnaire.

## 5. Summary

So far, it was highlighted the major steps that undertaken in constructing a new questionnaire. Researcher was able to clearly link the questionnaire items to the theoretical construct that intend to assess. It was crucial to invest the time and effort to ensure that the items adequately assess the construct of interest.

## 6.Validating a Questionnaire

### Initial validation

After the new questionnaire items pass through preliminary pilot testing and subsequent revisions, it was time to conduct a pilot test among the intended 60 respondents for initial validation. In this pilot test, the final version of the questionnaire is administered to a large representative sample of respondents for whom the questionnaire is intended.

### Reliability

The reliability of a questionnaire can be considered as the consistency of the survey results. As measurement error is present in content sampling, changes in respondents, and differences across raters, the consistency of a questionnaire can be evaluated using its internal consistency, test-retest reliability, and inter-rater reliability, respectively.

### Internal consistency

Internal consistency reflects the extent to which the questionnaire items are inter-correlated, or whether they are consistent in measurement of the same construct. Internal consistency is commonly estimated using the coefficient alpha, also known as Cronbach's alpha. Given a questionnaire *x*, with *k* number of items, alpha ($\alpha$) can be computed as:

$$\alpha = \frac{\kappa}{\kappa - 1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2}\right)$$

Where, $\sigma_i^2$ is the variance of item *i*, and $\sigma_x^2$ is the total variance of the questionnaire.

Cronbach's alpha ranges from 0 to 1 (when some items are negatively correlated with other items in the questionnaire, it is possible to have negative values of Cronbach's alpha). However, if a negative Cronbach's alpha is still obtained when all items are correctly scored, there are serious problems in the original design of the questionnaire, with higher values indicating that items are more strongly interrelated with one another. Cronbach's $\alpha = 0$ indicates no internal consistency (i.e., none of the items are correlated with one another), whereas $\alpha = 1$ reflects perfect internal consistency (i.e., all the items are perfectly correlated with one another). In this study Cronbach's alpha was calculated 0.840 with 33 items which has been suggested to indicate good internal consistency indicated in table no. 1

**Table no:1 Indicated the Cronbach's alpha value with 33 items.**

**Reliability Statistics**

| Cronbach's Alpha | N of Items |
|---|---|
| 0.840 | 33 |

IJASRM

International Journal of Advanced Scientific Research and Management, Volume 4 Issue 5, May 2019

www.ijasrm.com

ISSN 2455-6378

**Table no. 2: Item-Total Statistics**

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|
| Q1 | 53.0000 | 89.153 | .514 | .831 |
| Q2 | 53.0833 | 92.112 | .399 | .835 |
| Q3 | 53.0833 | 91.773 | .441 | .835 |
| Q4 | 52.6667 | 90.395 | .328 | .836 |
| Q5 | 53.0000 | 87.627 | .648 | .828 |
| Q6 | 52.4167 | 82.790 | .682 | .822 |
| Q7 | 52.2500 | 82.394 | .706 | .821 |
| Q8 | 52.7500 | 94.597 | .031 | .845 |
| Q9 | 52.8333 | 95.226 | .000 | .844 |
| Q10 | 51.8333 | 101.497 | -.272 | .873 |
| Q11 | 52.8333 | 91.497 | .299 | .837 |
| Q12 | 52.6667 | 83.446 | .651 | .824 |
| Q13 | 52.7500 | 94.258 | .054 | .844 |
| Q14 | 52.5000 | 88.898 | .401 | .834 |
| Q15 | 52.6667 | 94.633 | .031 | .845 |
| Q16 | 53.0833 | 92.112 | .399 | .835 |
| Q17 | 52.0000 | 90.000 | .357 | .835 |
| Q18 | 52.2500 | 80.699 | .678 | .821 |
| Q19 | 52.3333 | 80.904 | .839 | .817 |
| Q20 | 52.6667 | 92.938 | .271 | .838 |
| Q21 | 52.6667 | 88.192 | .418 | .833 |
| Q22 | 52.5833 | 95.671 | -.032 | .845 |
| Q23 | 52.8333 | 89.972 | .425 | .833 |
| Q24 | 52.4167 | 91.773 | .381 | .835 |
| Q25 | 52.5833 | 91.434 | .265 | .838 |
| Q26 | 52.7500 | 92.394 | .229 | .839 |
| Q27 | 52.6667 | 87.345 | .551 | .829 |
| Q28 | 52.5000 | 97.203 | -.170 | .847 |
| Q29 | 52.6667 | 87.175 | .564 | .829 |
| Q30 | 52.9167 | 88.891 | .695 | .829 |
| Q31 | 53.3333 | 95.650 | .000 | .841 |
| Q32 | 53.0833 | 90.925 | .544 | .833 |
| Q33 | 53.0000 | 89.322 | .679 | .830 |

Above marked items were excluded from the questionnaire based on statistic, Cronbach's alpha value reached to 0.890 with 28 items indicated in table no. 3.

**Table no. 3 Represent Cronbach's alpha with 28 items.**

**Reliability Statistics**

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|---|---|---|
| .890 | .890 | 28 |

## Test-retest reliability

Test-retest reliability refers to the extent to which individuals' responses to the questionnaire items remain relatively consistent across repeated administration of the same questionnaire or alternate questionnaire forms. Provided the same individuals were administered the same questionnaires twice (or more), test-retest reliability can be evaluated using Pearson's product moment correlation coefficient (Pearson's *r*) or the intraclass correlation coefficient.

Pearson's *r* between the two questionnaires' responses can be referred to as the coefficient of stability. A larger stability coefficient indicates stronger test-retest reliability, reflecting that measurement error of the questionnaire is less likely to be attributable to changes in the individuals' responses over time. The duration between time 1 and time 2 was not too short, so that individuals may remember their responses in time 1, which may overestimate the test-retest reliability.

## Validity

The validity of a questionnaire is determined by analysing whether the questionnaire measures what it is intended to measure. In other words, are the inferences and conclusions made based on the results of the questionnaire (i.e., test scores). Two major types of validity should be considered when validating a questionnaire: content validity and construct validity.

## Content Validity

Content validity refers to the extent to which the items in a questionnaire are representative of the entire theoretical construct the questionnaire is designed to assess. Although the construct of interest determines which items are written and/or selected in the questionnaire development, content validity of the questionnaire should be evaluated after the initial form of the questionnaire is available. The process of content validation is particularly crucial in the development of a new questionnaire.

A panel of experts who are familiar with the construct that the questionnaire is designed to measure should be tasked with evaluating the content validity of the questionnaire. The domain experts, judges whether the questionnaire items are adequately measuring the construct intended to assess, and whether the items are enough to measure the domain of interest. Nonetheless, as the process of content validation depends heavily on how well the panel of experts can assess the extent to which the construct of interest is operationalized, the selection of appropriate experts is crucial to ensure that content validity is evaluated adequately. Items to assess content validity include:

- The questions were clear and easy

- The questions covered all the problem areas
- You would like the use of this questionnaire for future assessments.
- The questionnaire lacks important questions.
- Some of the questions violate your privacy.

A concept that is related to content validity is face validity. Face validity refers to the degree to which the respondents or laypersons judge the questionnaire items to be valid. Such judgment is based less on the technical components of the questionnaire items, but rather on whether the items appear to be measuring a construct that is meaningful to the respondents.

## Construct validity

Construct validity is the most important concept in evaluating a questionnaire that is designed to measure a construct that is not directly observable (e.g., anxiety, impact, quality of life). If a questionnaire lacks construct validity, it will be difficult to interpret results from the questionnaire, and inferences cannot be drawn from questionnaire responses to a behaviour domain. The construct validity of a questionnaire can be evaluated by estimating its association with other variables (or measures of a construct) with which it should be correlated positively, negatively, or not at all. In practice, the questionnaire of interest is administered to the same groups of individuals. Correlation matrices are then used to examine the expected patterns of associations between different measures of the same construct, and those between a questionnaire of a construct and other constructs. It has been suggested that correlation coefficients of 0.1 should be considered as small, 0.3 as moderate, and 0.5 as large indicated in table no. 4.

**Table no. 4: Item-Total Statistics**

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|
| Q1 | 44.5000 | 94.831 | .484 | .886 |
| Q2 | 44.5833 | 97.535 | .398 | .888 |
| Q3 | 44.5833 | 96.857 | .478 | .887 |
| Q4 | 44.1667 | 96.582 | .272 | .890 |
| Q5 | 44.5000 | 93.305 | .614 | .883 |
| Q6 | 43.9167 | 87.027 | .736 | .878 |
| Q7 | 43.7500 | 86.631 | .760 | .878 |
| Q8 | 44.2500 | 99.852 | .047 | .895 |
| Q11 | 44.3333 | 96.328 | .345 | .888 |
| Q12 | 44.1667 | 88.446 | .660 | .881 |
| Q13 | 44.2500 | 99.343 | .081 | .895 |
| Q14 | 44.0000 | 93.729 | .434 | .887 |
| Q16 | 44.5833 | 97.535 | .398 | .888 |
| Q17 | 43.5000 | 95.339 | .358 | .888 |
| Q18 | 43.7500 | 85.784 | .677 | .880 |
| Q19 | 43.8333 | 85.056 | .898 | .874 |
| Q20 | 44.1667 | 99.124 | .191 | .891 |
| Q21 | 44.1667 | 94.379 | .363 | .889 |
| Q22 | 44.0833 | 100.417 | .032 | .894 |
| Q23 | 44.3333 | 95.480 | .413 | .887 |
| Q24 | 43.9167 | 97.027 | .397 | .888 |
| Q25 | 44.0833 | 96.518 | .287 | .890 |
| Q26 | 44.2500 | 97.987 | .216 | .891 |
| Q27 | 44.1667 | 91.836 | .608 | .883 |
| Q29 | 44.1667 | 92.175 | .583 | .883 |
| Q30 | 44.4167 | 93.806 | .738 | .882 |
| Q32 | 44.5833 | 96.349 | .539 | .886 |
| Q33 | 44.5000 | 95.169 | .621 | .884 |

## Sample size

Guidelines for the respondent-to-item ratio ranged from 2:1 (i.e., twenty respondents for a 10-item questionnaire). Others suggested that sample sizes of 50 should be considered as good,100 as very good, and 200 or more as excellent. Given the variation in the types of questionnaire being used, there are no absolute rules for the sample size needed to validate a questionnaire. As larger samples are always better than smaller samples, it is recommended that investigators utilize as large a sample size as possible.

## 7. Conclusion

In this study, Researcher have provided guidelines on how to develop and validate a questionnaire for the assessment of effectiveness of physical education curriculum for school children in obtaining the physical education objectives. The development of a questionnaire requires investigators' thorough consideration of issues relating to the format of the questionnaire and the

meaning and appropriateness of the items. Once the development stage is completed, it is important to conduct a pilot test to ensure that the items can be understood and correctly interpreted by the intended respondents. The validation stage is crucial to ensure that the questionnaire is psychometrically sound. Although developing a questionnaire is no easy task, the processes outlined in this article should enable researchers to end up with questionnaires that are efficient and effective in the target populations.

## References

[1] Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. BMJ. 2004; 328:1312–5. [PMC free article] [PubMed]
[2] Crocker L, Algina J. Introduction to Classical and Modern Test Theory. Mason, Ohio: Cengage Learning; 2008.
[3] Davis TC, Mayeaux EJ, Fredrickson D, Bocchini JA, Jr, Jackson RH, Murphy PW. Reading ability of parents compared with reading level of pediatric patient education materials. Pediatrics. 1994; 93:460–8. [PubMed]
[4] Bell A. Designing and testing questionnaires for children. J Res Nurs. 2007; 12:461–9.
[5] Wong DL, Baker CM. Pain in children: Comparison of assessment scales. Okla Nurse. 1988; 33:8. [PubMed]
[6] Stone E. Research Methods in Organizational Behavior. Glenview, IL: Scott Foresman; 1978.
[7] Hinkin TR. A brief tutorial on the development of measures for use in survey questionnaires. Organ Res Methods. 1998; 2:104–21.
[8] Harrison DA, McLaughlin ME. Cognitive processes in self-report responses: Tests of item context effects in work attitude measures. J Appl Psychol. 1993; 78:129–40. [PubMed]
[9] Price JL, Mueller CW. Handbook of Organizational Measurement. Marshfield, MA: Pitman; 1986.
[10] Harrison DA, McLaughlin ME. Exploring the Cognitive Processes Underlying Responses to Self-Report Instruments: Effects of Item Content on Work Attitude Measures. Academy of Management Annual Meetings. 1991:310–4.
[11] Embretson SE, Reise SP. Item Response Theory for Psychologists. Mahwah, N.J: Lawrence Erlbaum Associates, Publishers; 2000.
[12] Lindwall M, Barkoukis V, Grano C, Lucidi F, Raudsepp L, Liukkonen J, et al. Method effects: The problem with negatively versus positively keyed items. J Pers Assess. 2012; 94:196–204. [PubMed]
[13] Stansbury JP, Ried LD, Velozo CA. Unidimensionality and bandwidth in the Center for Epidemiologic Studies Depression (CES-D) Scale. J Pers Assess. 2006; 86:10–22. [PubMed]
[14] Tsang S, Salekin RT, Coffey CA, Cox J. A comparison of self-report measures of psychopathy among non-forensic samples using item response theory analyses. Psychol Assess. [In press] [PMC free article] [PubMed]
[15] Leung WC. How to design a questionnaire. Stud BMJ. 2001;9
[16] Artino AR, Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No 87. Med Teach. 2014; 36:463–74. [PMC free article] [PubMed]
[17] Schultz KS, Whitney DJ. Measurement Theory in Action: Case Studies and Exercises. Thousand Oaks, CA: Sage; 2005.
[18] Schmitt NW, Stults DM. Factors defined by negatively keyed items: The results of careless respondents? Appl Psychol Meas. 1985; 9:367–73.
[19] Thurstone LL. Multiple-Factor Analysis. Chicago, IL: University of Chicago Press; 1947.
[20] Churchill GA. A paradigm for developing better measures of marketing constructs. J Mark Res. 1979; 16:64–73.
[21] Perneger TV, Courvoisier DS, Hudelson PM, Gayet-Ageron A. Sample size for pre-tests of questionnaires. Qual Life Res. 2015; 24:147–51. [PubMed]
[22] Bowling A, Windsor J. The effects of question order and response-choice on self-rated health status in the English Longitudinal Study of Ageing (ELSA) J Epidemiol Community Health. 2008; 62:81–5. [PubMed]
[23] Lee S, Schwarz N. Question context and priming meaning of health: Effect on differences in self-rated health between Hispanics and non-Hispanic Whites. Am J Public Health. 2014; 104:179–85. [PMC free article] [PubMed]
[24] Schwarz N. Self-reports: How the questions shape the answers. Am Psychol. 1999; 54:93–105.
[25] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related

quality of life measures: Literature review and proposed guidelines. J Clin Epidemiology. 1993; 46:1417–32. [PubMed]
[26] Beaton D, Bombardier C, Guillemin F, Ferraz M. Recommendations for the Cross-Cultural Adaptation of the DASH and Quick DASH Outcome Measures. Toronto: Institute for Work and Health; 2007.
[27] Hendrickson WD, Russell IJ, Prihoda TJ, Jacobson JM, Rogan A, Bishop GD, et al. Development and initial validation of a dual-language English-Spanish format for the Arthritis Impact Measurement Scales. Arthritis Rheum. 1989; 32:1153–9. [PubMed]
[28] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25:3186–91. [PubMed]
[29] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
[30] Nunnally J. Psychometric Theory. New York: McGraw-Hill; 1978.
[31] Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. J Pers Assess. 2003;80:99–103. [PubMed]
[32] Wilkinson L the Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. Am Psychol. 1999;54:594–604.
[33] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
[34] Dawson B, Trapp RG. Basic and Clinical Biostatistics. 3rd ed. Norwalk, Conn: Lange Medical Books; 2001.
[35] Grootscholten C, Bajema IM, Florquin S, Steenbergen EJ, Peutz-Kootstra CJ, Goldschmeding R, et al. Inter-observer agreement of scoring of histopathological characteristics and classification of lupus nephritis. Nephrol Dial Transplant. 2008;23:223–30. [PubMed]
[36] Berry KJ, Mielke PW. A generalization of Cohen's kappa agreement measure to interval measurement and multiple raters. Educ Psychol Meas. 1988;48:921–33.
[37] Murphy KR, Davidshofer CO. Psychological Testing: Principles and Applications. Upper Saddle River, NJ: Prentice Hall; 2001.