9th FIEP European Congress
7th International Scientific Congress "Sport, Stress, Adaptation"

Physical Education and Sport – Competences for Life

9 -12 October 2014

National Sports Academy "Vassil Levski", Sofia, Bulgaria

CONGRESS PROCEEDINGS

SCIETIFIC JOURNAL
EXTRA ISSUE, 2014

WWW.FIEP2014SOFIA.COM

MOBAQ-LUX8 – A COMPETENCE-ORIENTED TEST BATTERY FOR 8 YEAR-OLD LUXEMBOURGISH STUDENTS: ITEM ANALYSIS

Claude Scheuer, Prof. Dr. Andreas Bund, Prof. Dr. Werner Becker
University of Luxembourg, Luxembourg
E-mail: claude.scheuer@uni.lu

Key words: *motor qualifications; motor test; test validation; item analysis*

INTRODUCTION

The University of Luxembourg is conducting the project of Basic Motor Qualifications (MOBAQ[1]) to gather more information about the motor status and potential deficiencies of students at the age of 8 years. The MOBAQ-concept is an innovative competence oriented concept to evaluate basic motor qualifications of students permitting PE teachers to foster students individually [2]. The aim of the project is the elaboration of competence-oriented test items according to the MOBAQ-approach [3] and according to quality criteria of standardized tests. To reach this goal, two pilot studies will be realized. The objective of the first study – of which a part of the results will be presented in this article – involves the generation of a psychometrically high quality item pool, which reflects the dimensions, facets and levels of the theoretical competence model (study I). For this purpose, statistical analysis of the developed tasks (item selection) using methods of classical test theory are performed. It is empirically tested whether the previously on a theoretical basis determined test dimensions are empirically valid. In a next step, the test battery is used in a pilot study (study II) to establish a diagnosis tool for pedagogical purposes in several school classes of the class level 3 in elementary school in Luxembourg. Those results should help to identify students with remediation needs on school and classroom level, in order to be able to suggest specific services and offers to students and their parents.

BACKGROUND OF THE STUDY

In opposition to usual approaches based on abilities and the statistical legitimation of minimal standards, the MOBAQ-approach fixes normatively basic motor qualifications as minimal requirements for students to be able to participate in the movement or sports culture. Thus, a basic motor qualification is a motor task that is (1) sufficiently complex; (2) related to a specific context, and (3) codified in a binary way (performed successfully vs. not performed successfully) [3, 4, 5, 6, 7]. This approach and the experiences made during the development of MOBAQ-NRW and MOBAQ-LUX12 have led to several criteria for MOBAQ-test-items [3]: (1) There is a consensus of the minimal requirements for children or adolescents to be able to participate in the culture of human motion in the sense of cultural participation; (2) the test situations do not demand specific technical requirements but are designed functionally in order to give place for individual solutions; (3) the limitations of the minimal standard are accepted as verisimilar for the living environment and have no time or other measurable limits or expectations; (4) all qualifications needed to solve a problem can principally be learned or reached by all children or adolescents, which means that their physical requirements are not relevant and that the needed learning places are accessible for everyone. Further relevant criteria are an easy-made planning and

[1]MOBAQ is the German acronym of the project and stands for "Motorische Basisqualifikationen".

organization of the tests, high movement intensity during test situations, the need of concentration to solve the test situation and enough time for practice before the tests. Examples of MOBAQ-test-items are described by Scheuer, Bund, and Becker (2013).

METHODS

Test dimensions and item collection

A development group consisting of associated teachers, students and scientists working in the MOBAQ-project designed the total of 29 items in 6 test dimensions. The test construction consists of 6 test dimensions: (1) Moving on equipment; (2) Moving in water; (3) Running and jumping; (4) Rolling and riding; (5) Playing with small devices; and (6) Playing with balls. Each test dimension counts five test items (tasks), as e. g. the test dimension "Moving on equipment": (1) Balancing; (2) Climbing; (3) Swinging; (4) Stabilizing and (5) Rotating. Figure 1 shows the different stages in the process of the item development (**Fig. 1**)

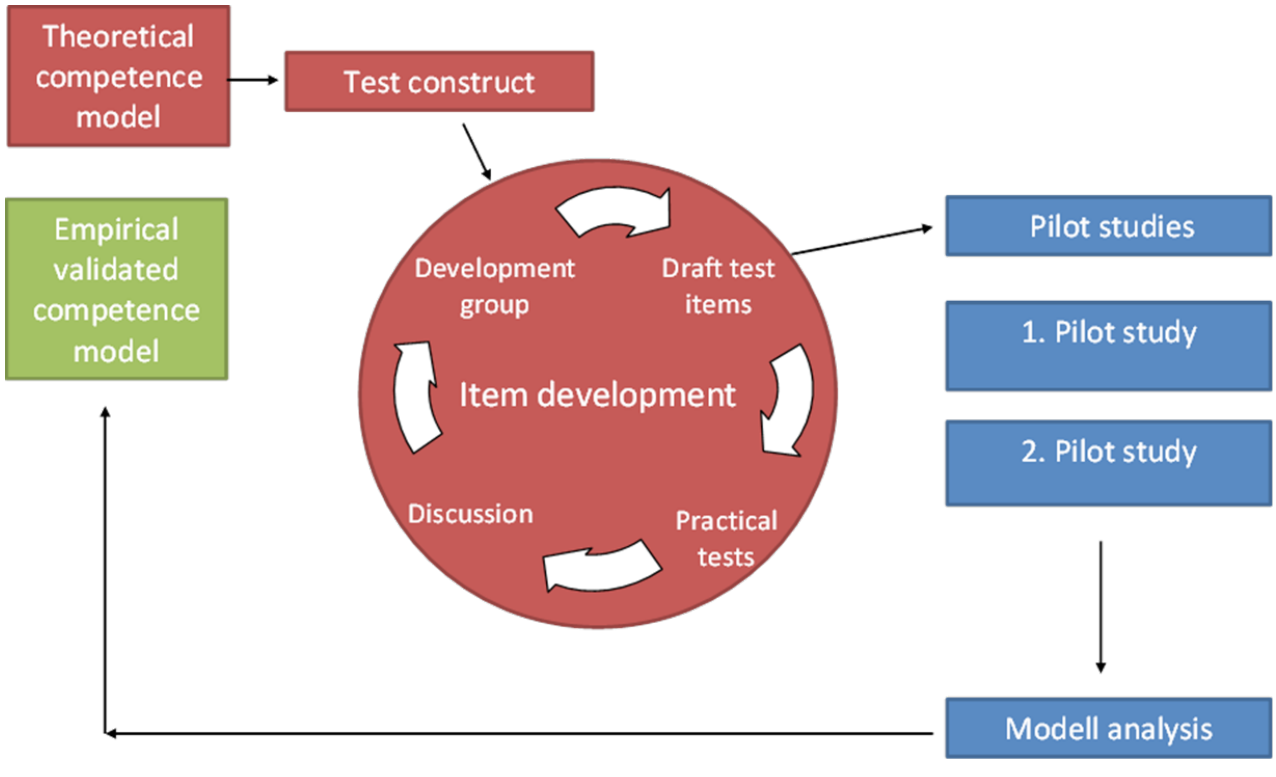

Fig. 1. Process of test and item development (Scheuer, Bund, & Becker, 2013, p. 344)

Validation of the MOBAQ-LUX8 test instrument

In study I a pre-test is conducted to validate MOBAQ-LUX8 scientifically in order to develop an adequate test design (including a test manual and test materials for teacher training). Test data of a sample of 113 students in eight classes allow to validate empirically the test construction, to select and revise the test items and to optimize the test instructions. The classical test criteria will be validated as described in table 1 below.

In study II, the impact of possibly relevant variables (e.g., gender, physical activity, social and economic background, migration background) will be taken into consideration, as well as in comparison to other tests. Furthermore, confirmatory factor analysis (CFA) and exploratory factor analysis (EFA) to revise the theoretical structure of the test battery will be conducted using the study data. The analysis of the data of study II is not subject of this article.

Sample

The review of the test items was performed on a sample of n=113 students from eight classes of class level 3 in Luxembourg.

Procedure

The tasks to capture the basic motor qualifications of 8-year-old students were tested in different dimensions according to the MOBAQ-approach based on the criteria for the development of MOBAQ-test-items. A pre-test for the scientific substantiation of MOBAQ-LUX8 in Luxembourg was conducted in order to develop a suitable test procedure.

Data analysis

The collected data were examined for the classical test criteria objectivity, reliability and validity. The findings provide the basis for the subsequent item selection and item re-construction. The aim was to have a test battery consisting of three test items per test dimension after the study I testing. Table 1 gives an overview on the evaluation of the different test criteria.

Table 1. Test Quality Criteria

Criteria	Procedure
Validity	Content validity: rating of the items by experts
	Construct validity: correlation analysis in the test dimensions to test the item characteristics
	Criterion validity: teacher rating / assignment in PE
Reliability	Re-test in one class
Objectivity	Standardization to assure objectivity in implementation, analysis and interpretation
	Rater compliance in assessment of video recording of min. 20 students

Objectivity. Standardisation. In order to ensure the objectivity of implementation, evaluation and interpretation, a written test manual with a description of the 29 test items affecting the test setup, the test procedure and the evaluation criteria for task assessment was compiled and given to the test administrators.

Inter-rater objectivity. To verify the objectivity, between nine and 51 students were filmed for each MOBAQ task depending on the test item. These recordings were presented to four PE teachers with a request to evaluate the subjects' task-coping. The ratings were then tested for compliance. Such coincidence checks at nominal scaled data can be made by defining Cohen's Kappa. Since Cohen's Kappa can only be calculated for the agreement between two raters, Fleiss` Kappa – representing the generalization to, specifically in this case, four ratings – had to be calculated. There were specified both Fleiss' Kappa values for the mean agreement on the contributor (Intra-class-correlation for mean of ratings), as well as the probability that a rater chooses the same category as any other rater (Intra-class-correlation for single rating).

Reliability. To check the reliability of MOBAQ, the test tasks were repeated in one class. By calculating the mean for each test dimension (i.e., summing the task values divided by the number of tasks) an interval-scaled value is obtained for which a test-retest-coefficient (r) can be determined. However, on task level a different approach was needed. Since the evaluation of tasks is carried out on nominal scale (performed successfully vs. not performed successfully), the consistency of the results was calculated using Cohen's Kappa.

Validity. Content validity. The content validity of the test items was examined by an expert rating of two physical education teacher and two sports scientists from the field of Physical Education.

Construct validity. A correlation analysis within the test dimensions was performed to verify the item assignment. Furthermore, we checked the overall model fit by calculating a selectivity coefficient.

Criterion validity. The school grade in physical education (from 1 = very good to 6 = very bad) given by the teachers was taken as external criterion which should be positively correlated with the MOBAQ-test performance in case of given validity.

RESULTS

In the following the results of the item analysis of study I are presented. The scale analysis will be produced in a further step.

Item analysis

Table 2. Results of the item analysis

Test item	ID	SD	Objectivity	Reliability	Validity	
			ICC	Cohen's Kappa	Construct	Criterion
BAG 1: Balancing	.98	.146	.817**	n. c.	.083	-.046
BAG 2: Climbing	.94	.247	1.000	.634	.437**	-.193*
BAG 3: Swinging	.96	.204	1.000	1.000	.405**	-.227*
BAG 4: Stabilizing	.89	.311	.986**	1.000	.304**	-.136
BAG 5: Rotating	.86	.349	.863**	1.000	.351**	-.063
BIW 1: Sliding	.93	.255	.948**	.328	.164	-.142
BIW 2: Diving	.89	.321	.925**	1.000	.116	-.077
BIW 3: Jumping into water	.98	.151	-.280	n. c.	.306**	-.027
BIW 4: Driving	.80	.399	n. c.	.857	.205	-.009
BIW 5: Floating	.89	.321	.577*	n. c.	.312**	.034
LUS 1: Persistent running	.64	.483	n. c.	1.000	.242*	-.261*
LUS 2: Coordinated running	.97	.180	.909**	.762	.290**	-.233*
LUS 3: Orientated running	.95	.229	.868**	n. c.	.180	-.066
LUS 4: Rhythmic jumping	.80	.401	.826**	.595	.224*	-.187
LUS 5: Coordinated jumping	.85	.363	.714**	n. c.	.017	-.049
RFG 1: Controlled riding	.93	.254	.891**	n. c.	-.102	.064
RFG 2: Changing track	.97	.164	.734**	n. c.	.379**	.049
RFG 3: Braking and stopping	.96	.200	.971**	n. c.	.120	.132
RFG 4: Controlled sliding 1	.92	.277	.111	n. c.	.187	.319**
RFG 5: Controlled sliding 2	.99	.117	-.750	n. c.	-.055	.015
SKG 1: Throwing in a target 1	.87	.337	.818**	.276	.117	-.055
SKG 2: Throwing in a target 2	.74	.440	.899**	.378	.254*	.012
SKG 3: Hitting in a target	.49	.503	.987**	.865	.003	-.327**
SKG 4: Controlling with a stick	.78	.413	.968**	n. c.	-.061	-.032
SPB 1: Throwing and catching 1	.96	.192	1.000	n. c.	.290**	-.135
SPB 2: Throwing and catching 2	.72	.449	.828**	.737	.323**	-.041
SPB 3: Controlled dribbling 1	.87	.342	.884**	.759	.268**	-.268**
SPB 4: Controlled dribbling 2	.79	.409	.878**	1.000	.433**	-.295**
SPB 5: Shooting and stopping	.86	.352	.945**	.587	.313**	-.029

Note. ID: Item difficulty; SD: Standard deviation; ICC: Intra Correlation Coefficient; n. c.: not calculated (as all results of at least one rater are the same); * p<0,05; **p<0,01

As **table 2** shows, the item difficulty of the different items lies between .49 for the most difficult item and .99 for the easiest item. Even if the MOBAQ concept focuses on minimal standards, and should thus a priori lead to high item passing quotes, several items – like e.g. BAG1, BAG3, BIW3, LUS2, RFG3, RFG5 and SPB1 – have very low item difficulties. The ICC-value giving an indication on the objectivity of the test-items is in general very high, except for BIW3, LUS1 and RFG4, which gives a positive indication about the quality of the standardization and the description of the test items. In the same way, Cohen's Kappa – giving an indication about the reliability – is in general high. For several items Cohen's Kappa has not been calculated, as at least at one test moment all the students passed the item successfully. Finally, the correlation values for the construct and criteria validity are less satisfying, as they are mostly indicating weak correlations with either the test areas or the external criterion of the grade in physical education.

DISCUSSION

In the following, the results of the review of the quality criteria for the individual tasks – as described below – are summarized and discussed.

Item difficulty

A test item can be considered as suitable difficult, if the item difficulty, or consist quote of a test item, lies between .80 and .95. Since MOBAQ tasks are minimum qualifications, the individual test tasks should also have high pass rates. Thus, if the item difficulty is under .80, the task can be seen as too difficult and should be more suitable for older children. If the item difficulty is above .95, the task can be regarded as too easy and should be more suitable for younger children.

Thus, the results show that especially the items LUS1 and SKG3 are too difficult and could be rejected or should be adapted. In the same logic, several other items (value 0 in the table) should also be adapted.

Objectivity

In the inter-rater objectivity should exist a high correlation, i.e. the respective cross-correlation coefficient ICC should be above r = .70.

The ICC values reflect in general a very high correlation between the different raters, except for the items BIW3 and RFG4, which means that only small improvements are needed for the standardization of the test items, which is mainly given by the description of the test items in the test manual.

Reliability

In the Intra-rater agreement should exist a high correlation, i.e. the respective correlation coefficient Cohen's kappa should be above r = .70.

The high intra-rater agreement values show that the results in the different test items are mainly the same on two test points, which is an important pre-requisite for the quality of the test items. Except the items constructed to evaluate skills in "Throwing" (SKG1 and SKG2) seem to lead to different results, which might be explained by the fact that the students have to solve the problem situations in these items several times to be successful (e.g. throw an object in a goal three times out of six trials).

Validity

In the construct validity, the selectivity coefficient according to Spearman should be at least medium, i.e. r = .30. However, as the item difficulties are generally high due to the concept of basic motor qualifications, it must be taken under consideration, that the calculated selectivity coefficients will be lower in general [1].

This means that, even if the results generally show low correlations between the isolated test items and the respective test areas, it cannot be concluded at this moment of the test review that most of the test items do not fit in the proposed construct. It can even be supposed that the test areas BAG and SPB consist of test items that fit rather good in the respective test areas, as the correlations are highly

significant. In the other test areas, this is only the case for a few of the test items.

In the criterion validity, the correlation between the grade in physical education and the test result should at least be low, i.e. r > .20, or here r < -.20, as the criterion is negatively scaled.

As can be seen in table 3, the correlation of the test items with the criterion "grade in physical education" is mostly very unsatisfying. This can be explained by the choice of the criterion, as it consists of a rather broad and vague subjective evaluation by teachers who are not specialized in the field of physical education but as elementary school teachers have to be generalists.

Table 3. Summary of the item analysis

Test item	ID	Objectivity	Reliability	Validity	
		ICC	Cohen's Kappa	Construct	Criterion
BAG 1: Balancing	0	+	n. c.	-	-
BAG 2: Climbing	+	+	0	0	0
BAG 3: Swinging	0	+	+	0	+
BAG 4: Stabilizing	+	+	+	0	0
BAG 5: Rotating	+	+	+	0	-
BIW 1: Sliding	+	+	-	-	0
BIW 2: Diving	+	+	+	-	-
BIW 3: Jumping into water	0	-	n. c.	0	-
BIW 4: Driving	+	n. c.	+	-	-
BIW 5: Floating	+	-	n. c.	0	-
LUS 1: Persistent running	-	n. c.	+	-	+
LUS 2: Coordinated running	0	+	+	-	+
LUS 3: Orientated running	+	+	n. c.	-	-
LUS 4: Rhythmic jumping	+	+	-	-	0
LUS 5: Coordinated jumping	+	+	n. c.	-	-
RFG 1: Controlled riding	+	+	n. c.	-	-
RFG 2: Changing track	0	+	n. c.	0	-
RFG 3: Braking and stopping	0	+	n. c.	-	-
RFG 4: Controlled sliding 1	+	-	n. c.	-	-
RFG 5: Controlled sliding 2	0	+	n. c.	-	-
SKG 1: Throwing in a target 1	+	+	-	-	-
SKG 2: Throwing in a target 2	0	+	-	-	-
SKG 3: Hitting in a target	-	+	+	-	+
SKG 4: Controlling with a stick	0	+	n. c.	-	-
SPB 1: Throwing and catching 1	0	+	n. c.	-	0
SPB 2: Throwing and catching 2	0	+	+	0	-
SPB 3: Controlled dribbling 1	+	+	+	-	+
SPB 4: Controlled dribbling 2	0	+	+	0	+
SPB 5: Shooting and stopping	+	+	-	0	-

Note. Item difficulty (ID): ID between .80 and .95: +, ID between .70 and .80 or > .95: 0, ID < .70: -; Objectivity: Intra Correlation Coefficient (ICC): r > .70: +, r between .60 and .70: 0, r < .60: -; Reliability (Cohen's Kappa): r > .70: +, r between .60 and .70: 0, r < .60: -; Construct validity: r > .50: +, r between .30 and .50: 0, r < .30: -; Criterion validity: r > .20: +, r between .10 und .20: 0, r < .10: -; n. c.: not calculated (as all results of at least one rater are the same)

CONCLUSION

As the results show, it is inevitable that some of the constructed tasks do not fulfil the requirements of standardized tests. Thus, these items have to be rejected or – in the case of minor defaults – they have to be adapted. With a view on the construct validity, it can be concluded at this point that the theoretic construct of MOBAQ has some failings, as the construct validity is generally very low, even if the concept of MOBAQ of minimal standards brings these low levels forward. Further conclusions about the construct validity can only be taken after the calculation of the scale analysis in a next step and the factor analysis of the data of the upcoming study II with the improved items. Besides this, it appears that the criterion "PE grades" chosen to review the criterion validity is not appropriate. This is why in study II should be collected further data, as e. g. the grade of physical activity during the week, as control variables in order to review the criterion validity.

REFERENCES

1. Bortz, J., & Döring, N. Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler [Research methods and evaluation for human and social scientists]. Springer: Heidelberg. 2006.
2. Département Ministériel des Sports [Ministerial Department of Sports]. Plan d'action national „Gesond iessen – méi bewegen". Extension du plan d'action national dans le domaine spécifique de l'activité physique [National action plan „Eat healthy – move more". Extension of the national action plan in the specific domain of physical activity]. Retrieved July 1st, 2014, from http://www.sport.public.lu/fr/sport-loisir/pan-extension/pan_intro.pdf. 2011.
3. Kurz, D., & Fritz, T. Motorische Basisqualifikationen von Kindern. Ergebnisse einer repäsentativen Untersuchung in Nordrhein-Westfalen. Abschlussbericht über das Forschungsprojekt MOBAQ I und II [Basic motor qualifications of children. Results of a representative study in North Rhine-Westphalia. Final report on the research project MOBAQ I and II]. Universität Bielefeld: Fakultät für Psychologie und Sportwissenschaft. 2007.
4. Kurz, D., Lindemann, U., Rethorst, S., Scheuer, C., Maatmann, H., & Schröer, T. Motorische Basisqualifikationen – ein pädagogisch sinnvoller Ansatz? [Basic motor qualifications – a pedagogical reasonable approach?]. Handout dvs - Sektion Sportpädagogik, Jahreskonferenz 2012. Magglingen. 2012a.
5. Kurz, D., Lindemann, U., Rethorst, S., Scheuer, C., Maatmann, H., & Schröer, T. MOBAQ-LUX12 Testaufgaben Pilotphase. Unveröffentlichter Testleitfaden [MOBAQ-LUX12 Test items pilot phase. Unpublished test guidelines]. Universität Bielefeld and Universität Luxemburg. 2012b.
6. Scheuer, C., Bund, A., & Becker, W. MOBAQ-LUX8 – Elaboration of competence-oriented test-items for 8year-old Luxembourgish students. FIEP-Bulletin, 83 (special edition), 2013, pp. 343-345.
7. Scheuer. C. Innovative concepts in Physical Education in Luxembourg. In M.-K. Chin & C. R. Edginton Physical Education and Health. Global Perspectives and Best Practice (pp. 281-296). Sagamore Publishing: Urbana, IL. 2014..