PERIODICA POLYTECHNICA SER. SOC. MAN. SCI. VOL. 10, NO. 1, PP. 11–20 (2002)

A RATING SCALE ANALYSIS OF FEATURES FOR MULTIMEDIA COURSEWARE

Katalin HIDVÉGI–MIKLÓS

Department of Technical Education
Budapest University of Technology and Economics
Vocational Training PhD Program
H–1111 Budapest, Egry J. Street 1, Hungary
Phone: +36-1-463-2116, Fax: +36-1-463-1697
e-mail: hikati@solaris.ymmf.hu

Received: September 20, 2001

Abstract

There is an effort also in pedagogy to conclude on basis of measured values. Usually the objects of a measurement are people, the instruments of the observations are items which work together to define a variable. The probabilistic measurement models estimate the position of both people and items on the line of the variable in the same way. So these models are suited for sorting items, too. According to this feature of these models, if the items concern courseware's characteristics, the estimated positions of them along the scale may consider also the sequence of the characteristics. In accordance with the thoughts above this study presents an attempt at developing an 'instrument' which examines the features of multimedia courseware.

Keywords: probabilistic measurement models, features of multimedia courseware, rating scale analysis.

1. Purpose and Research Questions

The purpose of this study was to examine a set of features in order to reveal their connections, relations and clear up how they form groups. Also the main issue of this investigation was to get to know how the probabilistic measurement model chosen works on these types of data basis, whether we can get results that are in accordance with our former experiences.

The research questions can be formulated as follows:

- Do the items work together to define a single variable?
- What are the less fitting items and why?
- Is there any other dimension in addition to the Rasch one?

2. Method

2.1. Instrumentation

A questionnaire constructed to be used in this study included a list of the 25 features and a section for other data. These were gender, age, qualification and habits of using computer. The features were defined by statements as follows (*Table 1*):

Table 1.

1. The production of the courseware should be cheap.
2. A path should be given to be tracked.
3. The colours should be used consistently.
4. The functioning should be safe.
5. The resource of information should be – if it is possible – visual.
6. It should be transformable, adaptable easily.
7. Can the navigation be complicated if we know exactly, where we are?
8. The text should be short like a sketch.
9. We should hear music continuously.
10. The structure of the software should be simple.
11. It should contain exercises and questions for control.
12. The screens should be varied, differently from each other.
13. It should be made by a staff (teacher, psychologist, graphic sound engineer, programmer etc.).
14. Verbal explanation should be together with figures, pictures.
15. The continuation should be depending on answers given to controlling questions at certain points.
16. Printing should be allowed from screens.
17. The explanations should be written.
18. The screens should be divided into parts with permanent functions.
19. The topic should be chosen very carefully in order to demand and employ multimedia instrumentation.
20. The use of the software should be simple, easy to learn.
21. It should be engaging, colouring.
22. Should we wander freely round it?
23. It should be like a book.
24. It should have a part, a summary for main notions of the elaborated topic.
25. Its machine use should be slight.

This set of features was selected considering the list of TEEM evaluation system [9] and findings of DELTA Project [2]. This list contains features that seem to be not very important (music, being like a book) or very important (simple use, engagement) for good quality.

The instrument used a stem of 'Please circle a number to indicate how important each of these features to be used in a multimedia courseware *for learning a new topic*.' It was followed by a scale for rating standards for each item: 1 = it is not necessary or it may be harmful; 2 = not so important, it can be left off; 3 = important; 4 = very important; 5 = indispensable, essential.

2.2. Sample

The population for this study consisted of students of the Technical College in Dunaújváros after completing a course on multimedia software and some lecturers working at several universities in the country. Due to missing or incomplete data, a total of 90 instruments were obtained.

Table 2.

GENDER

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
f	17	18.9	18.9	18.9
m	72	80.0	80.0	98.9
s	1	1.1	1.1	100.0
Total	90	100.0	100.0	

Respondents included 17 females (f) and 72 males (m). One person did not give the gender (s) (*Table* 2).

Table 3.

QUAL

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	3	3.3	3.3	3.3
1	5	5.6	5.6	8.9
2	82	91.1	91.1	100.0
Total	90	100.0	100.0	

Most of them (82) had secondary qualification (2), 5 persons graduated from a university (1) and 3 people gave no data as for the qualification (0) (*Table*3).

According to the age most of the respondents were young (1= under 20 years, 2= 20–30 years) and only 5 people were above 40 (4= 40–50 years, 5= 50–60 years, 6= more than 60) (*Table* 4).

Table 4.

AGE

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
1	20	22.2	22.2	22.2
2	65	72.2	72.2	94.4
4	3	3.3	3.3	97.8
5	1	1.1	1.1	98.9
6	1	1.1	1.1	100.0
Total	90	100.0	100.0	

Table 5.

HAB5

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	50	55.6	55.6	55.6
1	40	44.4	44.4	100.0
Total	90	100.0	100.0	

More than half of the respondents (50) have not used (0) educational software yet but most of them (48, according to other variables) use several computer programs regularly. The rest of people (40) filled the questionnaire with experience in using educational software (1) (*Table* 5).

2.3. Data Analysis

A rating scale analysis was performed on the data basis with the help of software MINISTEP v3.12. This program was developed in MESA Psychometric Laboratory, at Chicago University. First of all items were investigated according to their fitting and the most surprising responses were analysed in order to get to know the reason of misfitting. Omitted the 5 less fitting items the measure order was made for the rest of items. At the end a residual factor analysis was performed to find some other dimensions.

3. Findings

3.1. Item Fit Analysis

According to the fit statistic (*Table* 6) five items have OUTFIT MNSQ values much larger than 1.

Table 6.

Entry number	Raw score	Count	Measure	Error	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Score corr.		que
9	174	90	1.95	0.14	1.80	4.2	1.72	3.9	A	0.34	q9
4	400	90	−1.51	0.15	1.50	2.6	1.55	2.7	B	0.11	q4
25	284	90	0.23	0.11	1.42	2.8	1.44	2.9	C	0.17	q25
7	282	90	0.26	0.11	1.38	2.5	1.37	2.5	D	0.14	q7
13	326	90	−0.31	0.12	1.32	2.3	1.31	2.2	E	0.33	q13
11	325	90	−0.30	0.12	1.13	1.0	1.13	1.0	F	0.23	q11
8	240	90	0.83	0.12	1.06	0.4	1.07	0.4	G	0.41	q8
6	302	90	0.00	0.11	1.01	0.1	1.05	0.4	H	0.31	q6
22	328	90	−0.34	0.12	1.04	0.3	1.03	0.3	I	0.31	q22
24	319	90	−0.22	0.11	1.02	0.1	1.03	0.3	J	0.28	q24
1	280	90	−0.28	0.11	1.00	0.0	1.02	0.1	K	0.22	q1
15	269	90	0.43	0.12	0.96	−0.3	0.97	−0.2	L	0.30	q15
18	273	90	0.38	0.12	0.89	−0.8	0.88	−0.9	M	0.32	q18
12	248	90	0.72	0.12	0.88	−0.9	0.84	−1.2	l	0.47	q12
2	343	90	−0.54	0.12	0.85	−1.2	0.87	−1.0	k	0.28	q2
23	223	90	1.08	0.12	0.85	−1.1	0.80	−1.4	j	0.36	q23
16	350	90	−0.64	0.12	0.85	−1.3	0.84	−1.2	i	0.42	q16
19	328	90	−0.34	0.12	0.84	−1.3	0.83	−1.4	h	0.40	q19
21	276	90	0.34	0.11	0.82	−1.5	0.81	−1.5	g.	0.63	q21
10	340	90	−0.50	0.12	0.79	−1.8	0.79	−1.7	f	0.45	q10
5	335	90	−0.43	0.12	0.77	−1.9	0.79	−1.8	e	0.29	q5
17	319	90	−0.22	0.11	0.78	−1.9	0.77	−1.9	d	0.29	q17
20	377	90	−1.06	0.13	0.78	−1.7	0.78	−1.7	c	0.45	q20
3	289	90	0.17	0.11	0.77	−1.9	0.77	−1.9	b	0.44	q3
14	320	90	−0.23	0.11	0.67	−2.9	0.68	−2.8	a	0.35	q14
Mean	302.	90.	0.00	0.12	1.01	−0.2	1.01	−0.2			
S.D.	48.	0.	0.69	0.01	0.27	1.8	0.27	1.8			

So the most misfitting items are 9 ‘music continuously’, 4 ‘safe functioning’, 25 ‘slight machine’, 7 ‘complicated navigation with map’ and 13 ‘made by staff’. Item 9 with the largest fit statistic is also with the highest scale value. Item 4 has the lowest scale value and the others are in the middle along the line. It is expected that high scoring people prefer to give point 5 to item 4 and not even they give to item 9. In the middle of the scale in the case of items 25 and 7 the expected score values for high scoring people are 3 or 4.

In order to penetrate their poor fit to the model we should examine the patterns in the data matrix. *Table 7* shows the most unexpected response-patterns.

Table 7.

Most unexpected responses

ques	Measure	pers
		8421778341187547552 5463332173884442754377 833166
	high↘	3867281713001267833203043071616259780542541486881
4 q4	**– 1.51 B**	..3...333......33......3...1.........2...........
20 q20	–1.06 c	................................2................
16 q16	–0.64 i	..................2..............................
10 q10	–0.50 f	.....................2...........................
5 q5	–0.43 e	3.......2........2...............................
22 q22	–0.34 I	..................1............................5.
13 q13	**–0.31** E	.2.........2...............1..................5..
11 q11	–0.30 F	...2..5
14 q14	–0.23 a	..2..
17 q17	–0.22 d	....5...
24 q24	–0.22 J	.1...
6 q6	0.00 H	1......................................1.1.......
3 q3	0.17 b	......................................5..........
25 q25	**0.23 C**	11........................1......511.5.15...5....
7 q7	**0.26 D**	.........1.11.......1..11.....5.....1.....5......
1 q1	0.28 K	..........1........1.......5...5....1..1........5
21 q21	0.34 g	............................1....................
18 q18	0.38 M	.............1.......5...5....5..................
15 q15	0.43 L	1...........................1....5.5.............
12 q12	0.72 l	.....5...1...55...5..............................
8 q8	0.83 G	...........5..........5......5............4......
23 q23	1.08 j	15....5..
9 q9	**1.95 A**	55..5.4......4.............5......5..............
		low
		8421778341187547552254633321738844427543771833166
		3867281713001267833 0304307161625978054254 486881

In accordance with their measure items 7, 13 and 25 are in the middle of the scale. However, there are high scoring people (i.e. 83, 48) who gave them much less points than some low scoring people (i.e. 1, 18, 38). These two items do not fulfil the basic requirement of order that any item is solved with greater probability by a person with greater measure.

The easiest item is 4 but some high scoring people value it too low. The hardest item is 9 but many of respondents value it too high.

The reason of this inconsistency can be the poor item specification or that the item's meaning is contestable.

Due to the lower fitting of these 5 items they are to be omitted.

3.2. Item Measure Order

Subsequent to omitting the most misfitting items the fit statistics were improved (*Table 8*). Their expectation is 0 and variance is 1.

Table 8.

Number of items	25	25	20	20
Statistics	infit ZSTD	outfit ZSTD	infit ZSTD	outfit ZSTD
Mean	−0.2	−0.2	−0.1	−0.1
S. D.	1.8	1.8	1.0	1.1

The measure order of the better fitting items is as follows (*Table 9*).

Table 9.

Entry number	Raw score	Count	Measure	Error	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Score corr.	que
23	223	90	1.29	0.14	1.02	0.1	0.96	−0.3	0.36	q23
8	240	90	0.98	0.13	1.25	1.6	1.26	1.6	0.41	q8
12	248	90	0.85	0.13	1.02	0.1	0.97	−0.2	0.48	q12
15	269	90	0.51	0.12	1.11	0.8	1.11	0.7	0.31	q15
18	273	90	0.45	0.12	1.03	0.2	1.02	0.1	0.31	q18
21	276	90	0.41	0.12	0.90	−0.8	0.89	−0.8	0.65	q21
1	280	90	0.35	0.12	1.15	1.1	1.17	1.2	0.21	q1
3	289	90	0.22	0.12	0.83	−1.3	0.82	−1.4	0.48	q3
6	302	90	0.03	0.12	1.11	0.8	1.14	1.1	0.35	q6
17	319	90	−0.22	0.12	0.87	−1.1	0.87	−1.1	0.30	q17
24	319	90	−0.22	0.12	1.13	1.0	1.15	1.1	0.29	q24
14	320	90	−0.23	0.12	0.76	−2.1	0.77	−1.9	0.34	q14
11	325	90	−0.31	0.12	1.21	1.5	1.22	1.6	0.29	q11
19	328	90	−0.35	0.12	0.94	−0.5	0.93	−0.5	0.39	q19
22	328	90	−0.35	0.12	1.11	0.8	1.10	0.8	0.37	q22
5	335	90	−0.46	0.12	0.83	−1.4	0.84	−1.3	0.33	q5
10	340	90	−0.53	0.12	0.87	−1.0	0.88	−1.0	0.44	q10
2	343	90	−0.58	0.12	0.93	−0.5	0.96	−0.3	0.30	q2
16	350	90	−0.69	0.13	0.91	−0.7	0.92	−0.6	0.43	q16
20	377	90	−1.14	0.14	0.85	−1.1	0.86	−1.0	0.42	q20
Mean	304.	90.	0.00	0.12	0.99	−0.1	0.99	−0.1		
S.D.	39.	0.	0.60	0.00	0.14	1.0	0.14	1.1		

The hardest item is 23 ‘like a book’ and the easiest one is 20 ‘simple, easy to learn using’. This location along the scale means that item 23 did not get too high scores not even from the high scoring people consequently it seems to be not important for a good courseware. In the case of item 20, on the contrary, even the low scoring respondents gave quite high scores so it seems to be an important feature.

3.3. Analysis of Residuals

Having performed a principal component analysis of standardised residual correlation for questions the first factor explains 2.1 of 20 residual variance units and the measurement dimension explains 5.8 units of people variance.

This factor does not give a powerful, strong new dimension; it is not expected in the case of this feature’s set. However, there are contrasted items like A (23, ‘like a book’) and a, b, c, d (11 ‘control’, 24 ‘summary’, 20 ‘simple, easy to learn using’, 14 ‘also verbal explanation’). The difference between these types of features can not be considered an accident (*Table 10*).

4. Conclusions

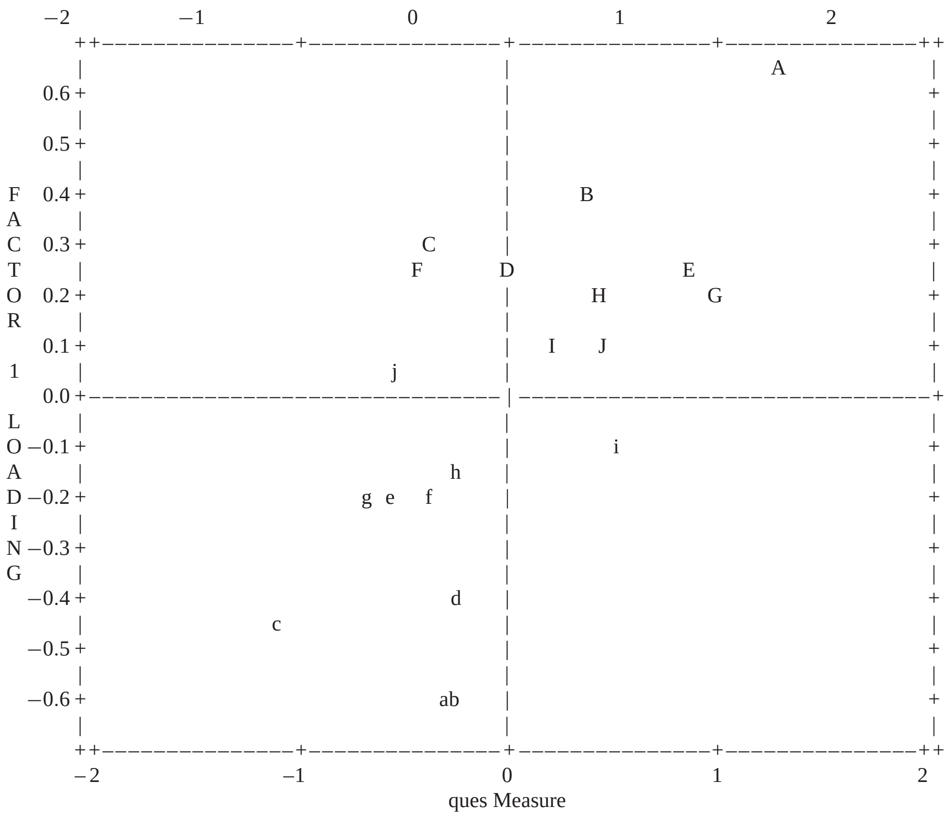
First of all, we need to ascertain which variable has been tried to measure. Perhaps it can be formulated as ‘competence for learning with computer software’. The ‘competence’ in general in pedagogy and also in this case is a complex notion having many strata i.e. experiences, knowledge, attitudes, motivation, desires, wishes, ideas. However, independently of its content it seems that these items (and persons) can define it properly by reason of results of this study. Of course, control and improvement are required.

The sequence of items according to their difficulty can be transformed into order of their importance. Easy items with high raw scores and low measures are important and hard items with low raw scores and high measures are not too important. At the ends of the scale there are items 23 ‘like a book’ as the hardest (measure 1.29) and item 20 ‘simple, easy to learn using’ as the easiest (measure -1.14). So according to this study the feature implied item 20 the most important for a good quality courseware. This conclusion is in accordance with findings of other investigations [2], [3] that emphasises the role of interface for quality and usability of educational software. Of course, the features ‘learnability’ and ‘simple use’ are related to interface closely.

This agreement does not verify the adequacy of whole sequence of importance but it is quite promising.

The analysis of residuals provided two groups of contrasted items: 23 ‘like a book’ and 11, 24, 20, 14 ‘control, summary, easy to learn using, also verbal explanation’. This second type of features can be interpreted as the expectation

Table 10.



laid against the new instruments in education. This sub-set of items contains claims like learning intensively by using more resources of information at the same time, handling it not taking away much energy. Users expect these new instruments to be different from the traditional ones.

This set of features examined in this study served the purpose of presentation of a mathematical instrumentation first of all. There are several feature-sets in the world used by several organisations for several aims (TEEM, OECD). It would be worth trying to analyse also other sets in order to get to know their inherent connections.

References

[1] ALKIN, M. C., Products for Improving Educational Evaluation. *Evaluation Comment*, V.2. (1970).

[2] BARKER, B. – KING, T., Evaluating Interactive Multimedia Courseware – a Methodology. Computers Ed. 4, 1993.

[3] GOBBATO, T. – MARCHIGIANI, E. – PARLANGELI, O. – BAGNARA, S., Experimental Evaluation of the Distance Delivery Course of Mathematics. MULTED Project n. 7524, University of Siena, 1996.

[4] IZSÓ, Lajos, Multimédia oktatási anyagok kidolgozásának és alkalmazásának pedagógiai, pszichológiai és ergonómiai alapjai. BME TK Budapest, 1998.

[5] ISO/IEC 9126 International Standard (1991).

[6] KÁRPÁTI, Andrea, Oktatási szoftverek minőségének vizsgálata, 2000, http://www.oki.hu/upsz.

[7] LEWY, A., Irányzatok az oktatási anyagok értékelésében. OOK, Veszprém, 1980, (manuscript).

[8] NAGY, Sándor, Az oktatástechnológia funkciója a pedagógiai integrációban. OOK–OTT, 1986.

[9] TEEM, 2000, http://www.teem.org.uk.

[10] TENNYSON, R. D., The Big Wrench vs. Integrated Approaches: The Great Media Debate. *Educational Technology Research and Development*, **3** (1994).

[11] VOSSEN, P. – MAGUIRE, M. – HEIM, J., Design Guide for Multimedia v2.1. INUSE, 1997.

[12] WRIGHT, B. D. – MASTERS, G. N., Rating Scale Analysis. Mesa Press, Chicago, 1982.