Validation of VARK learning modalities questionnaire using Rasch analysis

Validation of VARK learning modalities questionnaire using Rasch analysis

E D Fitkov-Norris and A Yeghiazarian
Kingston University, Kingston-upon-Thames KT2 7LB, UK

E-mail: e.fitkov-norris@kingston.ac.uk

Abstract. This article discusses the application of Rasch analysis to assess the internal validity of a four sub-scale VARK (Visual, Auditory, Read/Write and Kinaesthetic) learning styles instrument. The results from the analysis show that the Rasch model fits the majority of the VARK questionnaire data and the sample data support the internal validity of the four sub-constructs at 1% level of significance for all but one item. While this suggests that the instrument could potentially be used as a predictor for a person's learning preference orientation, further analysis is necessary to confirm the invariability of the instrument across different user groups across factors such as gender, age, educational and cultural background.

1. Introduction

A number of factors such as learner motivation, study skills and ability to assess their own learning needs have been identified as good predictors of student performance [1]. External factors such as classroom climate and aligning teaching style with learners' learning preferences may bring additional benefits for learners [2,3], with students being happier when taught using their preferred learning mode, namely Visual, Auditory, Read/Write or Kinaesthetic [4].

VARK type questionnaires have been used extensively to test the learning mode preference of postgraduate and undergraduate students on a number of degrees such as nursing, psychology and business [5,6]. Self-reported questionnaires are often used by educationalists to evaluate the propensity of learners to use particular learning approaches and tools in order to assess good and bad learning practices and identify adequate learning support mechanisms [7]. These instruments often rely on parametric statistical analysis to build measurement scales and to make overall recommendations, although as the data is often collected using Likhert or binary (Yes, No) scales [8-10], this manipulation may not be suitable. Rash analysis can be used to transform the ordinal or binary response scales into linear interval scales that can be manipulated freely [11]. The analysis can also be used to identify items that do not fit the scale as well as for construct validation and development particularly when construct invariance across different groups of users (for example male and female or younger and older) is required [12]. Furthermore the analysis can be used to analyse the psychometric properties of a scale and identify overlapping response categories that can be merged [13].

This paper will use Rasch analysis to examine the internal construct validity of a questionnaire designed by Neil Fleming [14] which assesses the preferences for students to receive and process information using Visual, Auditory, Read/Write and Kinaesthetic modes. It consists of 16 scenarios each asking the respondent to identify all of the information processing modes they would adopt in a particular scenario. The number of ticks given for each category are counted to give an overall score for each learner's preference for this mode of information processing. Based on these scores the overall preference of the learner is identified by using normative [15] comparison to the average scores obtained by subjects who have completed the questionnaire in the past and recommendations for the best mode for learning and information processing are given. Although the VARK questionnaire has been analysed in the past using factor analysis, taking into account the testlet nature of the questions led to an improvement in the factor model fit. However the overall fit to the four learning modalities was still not statistically satisfactory which cast some doubt on its validity [16]. Furthermore, the suitability of the questionnaire for measuring individual learning preferences by combining its scales and computing averages from the binary responses has never been examined. Rasch analysis will be used to carry out a preliminary examination to assess the internal validity of the four sub-scales of the questionnaire and check if response patterns fit the interval scale for each of the four learning preference trait. Response patterns fitting the Rasch model well would be suitable for parametric statistical analysis and increase the confidence in the recommendations made by the instrument.

Rasch analysis has been used extensively for analysing both questionnaire and construct validity and reliability in a number of aspects of healthcare [11,17] and social sciences [18-20]. In the context of education and learning the most prevalent use of Rasch analysis is in assessing the match of question difficulty to the ability of students in multiple choice tests [21]. Only relatively recently has Rasch analysis been used to validate learning inventories [12], motivation [22-24] and learning preference constructs [25], although its use has been limited.

Rasch analysis tests the fit of responses to a questionnaire to a formal scale model developed by Georg Rasch which gives the expected responses if an interval scale measurement is to be achieved [11]. The model is a probabilistic form of Gutman scaling [26] which assumes strict ordering of the questions, from ones easiest for participants to agree with up to ones that are hardest to agree with, thus differentiating between participants with low and high preference for the latent trait being measured. The model can be applied to measuring instruments with both binary and polythomous items [18] and allows questionnaire items to be summed in order to provide a total score as an overall measure for the latent trait under examination. The binary form of the models takes the form of equation (1);

$$P_i\left(U_j = 1 \middle| \theta\right) = \frac{e^{\left(\theta - b_j\right)}}{1 + e^{\left(\theta - b_j\right)}} \qquad (1)$$

which expresses the probability of person i whose propensity to agree with the statements of the questionnaire is parameterised by latent trait θ agreeing with item j $\left(U_j = 1\right)$. The parameter b_j refers to item difficulty/ease of agreement.

2. Methodology

The VARK questionnaire was administered to 107 postgraduate students studying for a master's degree in business management or closely related subjects e.g. marketing and human resource management. The majority of students who completed the survey were female (68.2%). The average age of the participants was 26.8 years with a standard deviation 5.24 years. 81% of the students in the sample were international and came mainly from Europe, Asia and Africa. The frequency with which the students chose each of the four modes is shown in table 1. The most prevalent learning preferences

for this sample are Auditory and Kinaesthetic (t=1.32, df=100, two tailed p=0.187). Read/Write preference received significantly fewer ticks (t=-2.36, df=101, two tailed p=0.004) with Visual being the least popular (t=-6.22, df=97, two tailed p=0.000). The learning style preferences of business students classified by the VARK assessment tool shows that a large proportion of students (38.1%) did not have a particular preference and use all four modes. Over half of the students (58.1%) use a unimodal approach to learning, and a small minority used bimodal. The sample contained no students who used three modes. Of the 61 students who preferred a unimodal approach, the majority (46%) used Auditory, while the remaining three modes, Kinaesthetic, Visual and Read/Write, were spread equally among students (with approximately 18% expressing preference for each mode).

The analysis was carried out using IBM SPSS Statistics version 21.0 using the extended Rasch Model add-on eRm R package (http://www.01.ibm.com/support/docview.wss?uid=swg21488442). The testlet nature of the questionnaire suggests that to assess the questionnaire's overall structure validity a multilevel Rasch model needs to be considered [27]. However, multilevel models require a minimum of 200 observations to ensure that the parameter estimates are reliable [28]. Given the limitation of the small sample size, this paper reports on preliminary analysis that only considers the model fit of the four separate sub-scales of the VARK questionnaire.

Table 1. Number of ticks for VARK preference modes

	Mean	**Standard deviation**
Visual	4.44	2.51
Auditory	6.56	3.03
Read/Write	5.42	2.79
Kinaesthetic	6.17	2.72

3. Results

The results from the Rasch analysis on each of the four sub-scales of the VARK questionnaire are summarised below.

3.1 Visual

Overall the responses to Visual preference questions fit the Rasch model well, although the fit of VQ11 is affected by some outliers indicated by relatively high Outfit MSQ value and a statistical significance of p=0.003. The infit and outfit mean square values for all Visual items are in the range of 0.73-1.43, which is within the acceptable region of 0.5 - 1.5. The person fit responses are acceptable at 5% level of significance on all reported statistics. The person/item map (Figure 1(a)) shows that the items are mainly clustered to the right of the scale, suggesting that only people with stronger Visual modal preference are likely to agree with the items.

3.2 Auditory

Overall the responses to Auditory type items fit the Rasch model well, although the fit of AQ05 is affected by outliers indicated by a relatively high Outfit MSQ value and a statistical significance of p=0.013. The Auditory item infit and outfit mean square values are in the range of 0.64 - 1.32 which is acceptable. The person fit statistics are only acceptable at 1% level of significance for the Casewise Deviance, although this statistic is found to frequently make type 1 error [29]. The person/item map (Figure 1(b)) shows that the items are mainly clustered to the centre and the right of the scale, suggesting that only persons with moderate and stronger Auditory modal preference are likely to agree with the items.

3.3 Read/Write

Overall the responses to Read/Write items fit the Rasch model well, although the fit of RQ01 and RQ08 are affected by some outliers as indicated by a relatively high Outfit MSQ value and a statistical significance of p=0.036 and p=0.045 respectively. The Read/Write item infit and outfit mean square values are in the range of 0.74 – 1.25 and are acceptable. The person fit responses are only acceptable at the 1% level of significance for the Hosmer-Lemershow statistic, although the reliability of this test is affected by small sample sizes [29]. The person/item map (Figure 1 (c)) shows that the items are mainly clustered to the right of the scale, suggesting that only persons with stronger Read/Write modal preference are likely to agree with the items.

3.4 Kinaesthetic

Overall the responses to Kinaesthetic items fit the Rasch model well, although question KQ08 has higher MSQ and fails the significance testing at 5% with p=0.31. Overall the Read/Write item infit and outfit mean square values range between 0.78 – 1.26 and are acceptable. Although the person fit responses are rejected at 1% level of significance when using the Casewise Deviance, this statistic frequently makes type 1 error [29]. The person/item map show that the items are mainly clustered to the right of the scale, suggesting that only persons with stronger Kinaesthetic modal preference are likely to agree with the items.

4. Discussion

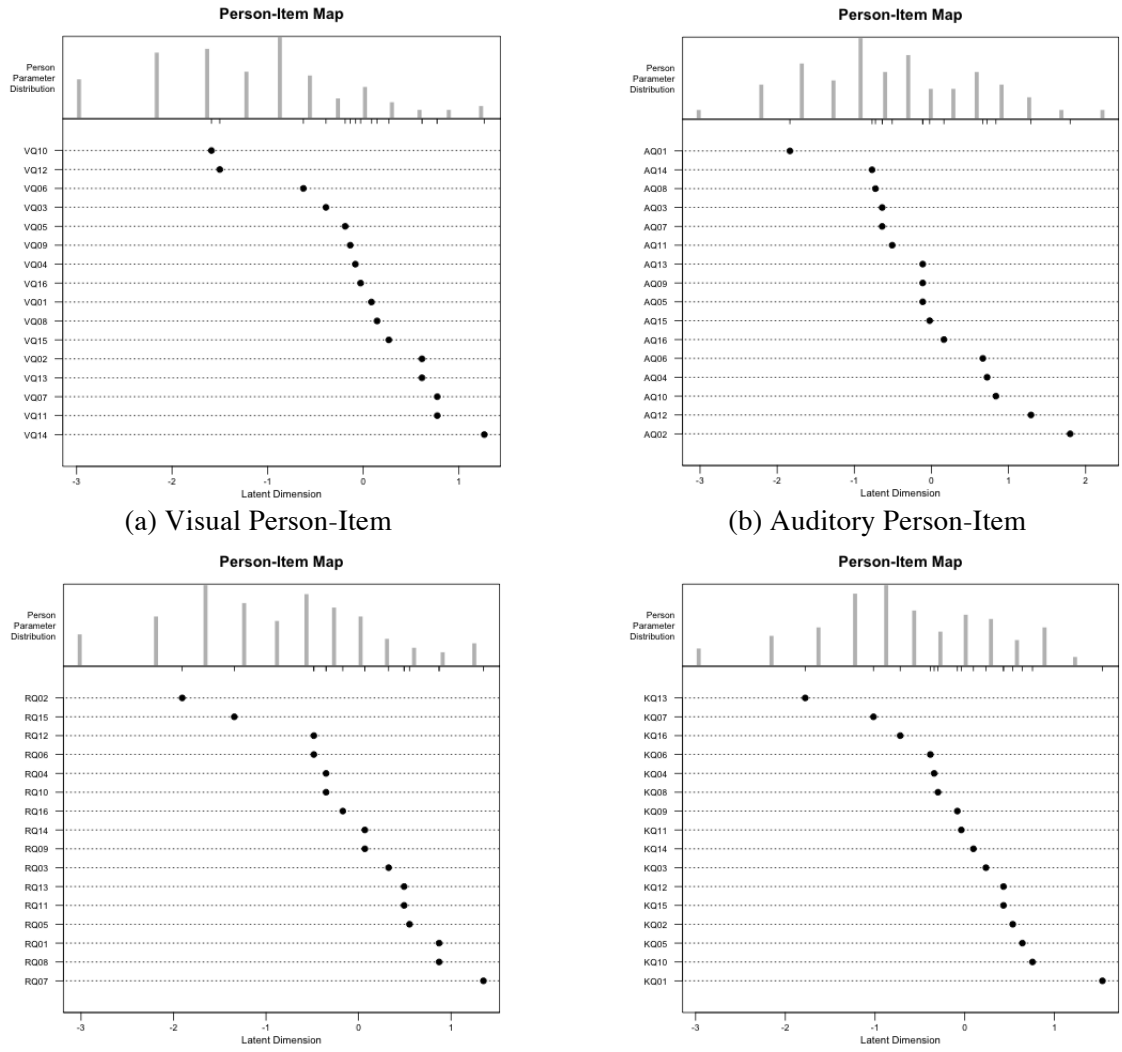
The results from the analysis confirm that overall the Rasch model fits the VARK questionnaire responses, and the data supports the internal validity of the four sub-constructs. The majority of items in the four sub-constructs, Visual, Auditory, Read/Write and Kinaesthetic preferences for learning match the data at 5% level of significance, and could be used as predictors for a person's learning preference orientation. The analysis confirmed that the instrument responses match the Rasch model and are, therefore, suitable for parametric statistical analysis. In addition, the binary nature of the responses would not have a negative impact on the reliability of any recommendations made to learners regarding their learning style preference. However, one Visual question had significance level less than 1% and four other questions had significance levels of less than 5% (see table 2).

Table 2. Scenario responses that do not fit the Rasch model

Scenario:	Problem response:	Preference:
Other than price, what would most influence your decision to buy a new non-fiction book?	The way it looks is appealing.	Visual
A group of tourists wants to learn about the parks...	Talk about, or arrange a talk for them about parks or wildlife reserves.	Auditory
You are helping someone who wants to go to your airport, town centre or railway station…	Write down the directions.	Read/Write
You have a problem with your heart. You would prefer that the doctor…	Gave you something to read to explain what was wrong.	Read/Write
	Used a plastic model to show what was wrong	Kinaesthetic

As mentioned, the questionnaire consists of 16 testlet scenarios each with four options, one for each type of learning preference. Previous research has indicated that the scenarios may introduce a bias in the answers within each testlet [16, 27]. Noting the specific scenarios that contain responses that do not fit the Rasch model, one could see that the Auditory and Read/Write options require the respondent to go to either considerably more effort to organise a talk or be in possession of a pen. Similarly, the Visual questions may clash with users' preference for value for money which is another option within the scenario, or external factors such as preference in younger users to read online rather

than purchase books. Two of the answers in scenario 8 do not match the expected item fit and this may be due to people choosing answers based on expectations from the scenario (visiting a doctor) rather than reflecting their true learning preference. The testlet nature of the model may also explain the small degree of overfit existing across some items as the responses to questions may be influenced by the particular scenario overarching the question.



(a) Visual Person-Item (b) Auditory Person-Item

(c) Read/Write Person-Item (d) Kinaesthetic Person-Item

Figure 1. Person-Item map results

Some person fit statistics also did not satisfy the 5% peel of statistical significance although this applied only for the Hosmer-Lemeshow and Casewise Deviance measures which are very sensitive to small sample size or much more likely to reject the null hypothesis, when it is in fact true [29]. The most robust of all item fit statistics, Collapsed Deviance, confirms item fit for all four sub-scales. The person-item figures confirm that in most cases people with moderate strength in a particular learning preference are likely to agree with the statements. Very few questions are likely to be agreed upon by people with low preference and, similarly, there are no questions that are likely to be agreed upon only by people with very strong learning preference. This is not entirely surprising for people with low learning preference as the questionnaire is designed to identify a positive preference for a learning approach, rather than dislike. However, the limited number of questions that identify strong preference

may be limiting the discriminatory power of the instrument. Due to the very small sample size it is not possible to guarantee its representativeness across all learning preference categories or to check the invariability of the instrument across different of groups of users across factors such as gender, age and background.

5. Conclusions and Recommendations

This pilot study assessed the internal validity of the four sub-scales of the VARK learning preference questionnaire and showed that the response patterns fit the Rasch model. The findings lend support to the questionnaire's suitability and reliability as an instrument for measuring learners' preferences for receiving and processing information in Visual, Auditory, Read/Write and Kinaesthetic ways. However, the limited sample size precluded the examination of the questionnaire's multi-level structure and thus further analysis using a significantly larger sample is required, before any recommendations are made for amending/removing items. In addition, a larger sample size will allow research into the invariability of the instrument across different groups of users grouped by gender, age and background, and provide basis for drawing conclusions about the overall suitability of the VARK measurement instrument.

References

[1] Credé M and Kuncel N R 2008 *Perspect Psychol Sci* **3** 425–53
[2] Popescu E 2009 *Advances in Web Based Learning–ICWL 2009* 343–52
[3] Hsieh C T, Mache M and Knudson D 2012 *Sport Biomech* **11** 108–19
[4] Haq S M, Yasmeen S, Ali S and Gallam F I 2012 *JRMC* **16** 191–3
[5] Dobson J L 2010 *Adv Physiol Educ* **34** 197–204
[6] Fitkov-Norris E D and Yeghiazarian A 2013 *12th ECRM for Business and Management Studies* (Guimaraes, Portugal) pp 144–52
[7] Coffield F, Moseley D, Hall E and Ecclestone K 2004 *Learning Styles and Pedagogy in Post-16 Learning* (Learning and Skills Research Centre)
[8] Brown W F and Holtzman W H 1955 *J Educ Psychol* **46** 75
[9] Entwistle N J, Thompson J and Wilson J D 1974 *High Educ* **3** 379–96
[10] Nonis S A and Hudson G I 2010 *J Educ Bus* **85** 229–38
[11] Tennant A and Conaghan P G 2007 *Arthritis Rheum* **57** 1358–62
[12] Waugh R F and Addison P A 1998 *Brit J Educ Psychol* **68** 95–112
[13] Pallant J F and Tennant A 2010 *Brit J Clin Psychol* **46** 1–18
[14] Fleming N D 1995 Proceedings of the 1995 *Annual HERSA Conference* **18** pp 308–13
[15] Boverman D M 1962 *Psychol Rev* 69 295–305
[16] Leite W L, Svinicki M and Shi Y 2010 *Educ Psychol Meas* **70** 323–39
[17] Belvedere S L and de Morton N A 2010 *J Clin Epidemiol* **63** 1287–97
[18] Kahler C, Strong D, Read J, Palfai T and Wood M 2004 *Psychol Addict Behav* **18** 322–33
[19] Hagquist C and Andrich D 2004 *Pers Indiv Differ* **36** 955–68
[20] Raudenbush S W, Johnson C and Sampson R J 2003 *Sociol Methodol* **33** 169–211
[21] Divgi D R 1986 *J Educ Meas* **23** 283–98
[22] Waugh R F 2002 *Brit J Educ Psychol* **72** 65–86
[23] Waugh R F 2003 *J Appl Meas* **4** 164–80
[24] Waugh R F 2002 *Brit J Educ Psychol* **72** 573–604
[25] Di Nisio R 2010 *Procedia Soc Behav Sci* **9** 373–7
[26] Andrich D 1985 *Sociol Methodol* **15** 33–80
[27] Wang W-C, Chng Y-Y and Wilson M 2005 *Educ Psychol Meas* **65** 5–27
[28] Brune K D 2011 *An Evaluation of Item Difficulty and Person Ability Estimation using the Multilevel Measurement Model with Short Tests and Small Sample Sizes,* PhD Thesis.
[29] Reise S P, Bentler P M and Mair P 2008 IRT Goodness-of-Fit Using Approaches from Logistic Regression *available at: http://escholarship.org/uc/item/1m46j62q*