

Evaluating the impact of a Quasi-ipsative scoring approach on the scoring of a VARK style questionnaire

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 J. Phys.: Conf. Ser. 772 012013

(<http://iopscience.iop.org/1742-6596/772/1/012013>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 141.241.26.77

This content was downloaded on 13/04/2017 at 09:00

Please note that [terms and conditions apply](#).

Evaluating the impact of a Quasi-ipsative scoring approach on the scoring of a VARK style questionnaire

E Fitkov-Norris and A Yeghiazarian

Kingston Business School, Kingston University, Kingston-upon-Thames, UK.

Email: e.fitkov-norris@kingston.ac.uk

Abstract. This paper demonstrated the application of a quasi-ipsative scoring approach to assess the relative strengths of individual preferences in a VARK style questionnaire. The approach identified a chi-squared test as more suitable method for analysing the type of data gathered by the VARK questionnaire. The results suggest that the quasi-ipsative chi-squared based approach does not appear to be as sensitive as the original t-test approach in identifying significant modalities. In order to increase the sensitivity of the test, the requirement for overall test significance has to be relaxed and individual cells considered in the analysis. The findings also put some doubt on the statistical validity of the original t-test approach as it also uses the deviation from the means (in the form of standard deviation) rather than statistical significance, as a tool for assessing the strength of individual preferences. The causes of the discrepancies between the two scorings techniques need to be examined further, before recommendations can be made on which approach is better suited to identifying the strength of individual preference for information input modality.

1. Introduction

This article assesses the impact of a new, quasi-ipsative scoring mechanism on the recommendations made by a multidimensional instrument measuring the strength of individual's learning style preference utilising dichotomous questions. The instrument, designed by Fleming [1] evaluates an individual's preference for Visual, Auditory, Read/Write and Kinaesthetic (VARK) mode of information input with the objective of identifying dominant traits. The instrument consists of 16 single stimuli questions, each presenting a different scenario and asks the user to tick all options that apply to them in a particular scenario. For example, a user with a single preference is expected to tick one option within each scenario, whereas a user with multimodal preferences is expected to tick several options within that scenario. The output from this scoring scheme is a 4 by 16 binary matrix indicating a user's preference for each of the 4 dimensions and the strength of each preference is identified by the frequency with which a particular dimension is chosen. Allowing users to choose the number of ticks that apply in a particular scenario fits relatively closely with an intuitive understanding of the way preferences are defined, namely, "Relatively stable evaluative judgments in the sense of liking or disliking a stimulus, or preferring it or not over other objects or stimuli ..." [2, p. 9]. The larger the number of ticks on each dimension, the stronger the user preference. Asking the users to tick all that apply also allows the instrument the flexibility to assess users with varying number of dominant preferences (unimodal, bimodal, tri-modal and multimodal).

However, this scoring technique also has some disadvantages. Firstly, different users can tick a different number of ticks in total. As a result, it is not possible to compare individual responses to the



population of all users, as the expected number of ticks in the population cannot be determined. One solution is to compare the number of ticks for a user in each dimension to the expected number of ticks in the sub-population of users who ticked the same number of ticks, see for example [3]. So if a user has ticked a total of 16 ticks, one can compare the number of actual ticks in each dimension to the expected number of ticks in each dimension for all users who ticked 16 options in total. Similarly, a user with 17 ticks in total will be compared to the expected values sub-population of users who chose 17 options in total. By definition the relative number of ticks in each dimension is expected to monotonically increase as the total number of ticks given by the user increase. Violation of this property can lead to discontinuities in the assessment of preferences, as each additional tick changes the reference point for the individual and the respective expected values. In some cases an extra tick could also lead to a significantly different recommendation on the person's preferences as the mean expected number of ticks for a preference can go down despite the fact that the overall number of ticks has gone up. For example, in Table 1 individual A had a total of 8 ticks in category R and was categorised as strong R, while individual B, with the same number of ticks in that category was categorised as mild R, although their profile is almost identical. Similarly, individual C is classified as having two preferences (AR), while individual D has a very similar profile, but is classified as having all 4 (VARK). In some cases, adding an extra tick can lead to loss of preferences (for example, individual E is tri-modal while individual F has a mild single preference).

Table 1 Examples of original VARK scoring

	V	A	R	K	Total ticks	Original scoring
Individual A	0	3	8	4	15	strong R
Individual B	0	3	8	5	16	mild R
Individual C	2	4	5	2	13	VARK
Individual D	2	4	5	3	14	VARK
Individual E	2	5	4	6	17	ARK
Individual F	2	5	4	7	18	mild K

Secondly, the technique compares the average number of tick in the respective sub population to the number of ticks the user selected using a t-test and uses the number of standard deviations away from the mean the users score to determine the strength of their preference. The t-test is designed to be used with quantitative data, which has interval or ratio measurement scale. However, the data gathered by the questionnaire is binary, confirming the presence of a preference but not providing any information on the relative strength of that preference. The analysis used makes the implicit assumption that each tick (across the 16 scenarios) denotes equal difference in preference from any options that have not been ticked. So an implicit assumption is made that a user ticking Auditory and Kinaesthetic in a particulate scenario has equal preferences for the two and in addition, the difference in preference between the Auditory and Kinaesthetic preferences selected and the Visual and Read/Write preferences that were not selected are also the same. This assumption has not been tested explicitly and therefore, using the t-test in the way may be inappropriate.

One solution to this problem is to change the setup of the questionnaire to a fixed-sum ranking question, where we require users to provide a ranking or relative order of their four preferences within each scenario (1st, 2nd, 3rd and 4th), for example. This approach will provide more detailed and accurate information on the actual preferences the user has and can be analysed using non-parametric tests such as Friedman's Test to compare ranks given across the 4 dimensions. However, the approach will increase the burden for the tests users and may lead to underside response behavior [4]. In addition, fixed-sum questions can lead to an artificial reduction on the response range and distorted scale relationships [5]. Although some solutions to the problem of fixed-sum ipsative scales have been pro-

posed [6], the suggested changes may lead to significant changes in the scales of the questionnaire and their introduction is not recommended without carrying to extensive testing and validation.

To ensure that the integrity of the existing questionnaire is not violated [3], this article proposes an alternative approach for the analysis of the existing scores using Chi-squared non-parametric test for association. The chi-square test is designed to compare frequencies of nominal data and does not make any implicit assumptions about the rating scale of the data. However, the total number of ticks given by each participant is different as the design of the questionnaire allows each question to carry up to 4 marks (ticks). As one of the assumptions of chi-squared test for association is that each question can carry only one count towards the total, this test cannot be used for normative comparison with the general responses within the population. To ensure that this assumption is not violated, the test has to be applied at the level of the individual. The test will compare the proportion of ticks (out of 16) the user has given to each dimension and can be used to identify the relative importance of each dimension within the context of the individual. This is a quasi-ipsative approach to personality measurement as the VARK instrument does not meet the fixed-sum prerequisite of strictly ipsative approaches [7]. Despite their relatively recent introduction, meta-analysis into personality trait measurement instruments suggests that quasi-ipsative approaches provide a more robust and reliable estimation of the trait compared to instruments adopting normative and purely ipsative approaches [7]. Therefore, this article will suggest a methodology for using Chi-squared test to assess the information processing preferences of users for visual, auditory, read/write and kinaesthetic information and will compare the efficacy of the quasi - ipsative approach to the classification results using the existing scoring mechanism.

2. Description of the method

The analysis was carried out on sample of 51 observations of VARK data provided by Neil Fleming and collected through the VARK online questionnaire version 7.1. (<http://vark-learn.com/the-vark-questionnaire/>). 73% of the sample was female, 23% male and 4% did not specify. 78% of respondents were students and 14% teachers. 45% of students were under 25 years old. Over half of respondents (57%) were from USA and next largest group was from UK (21%).

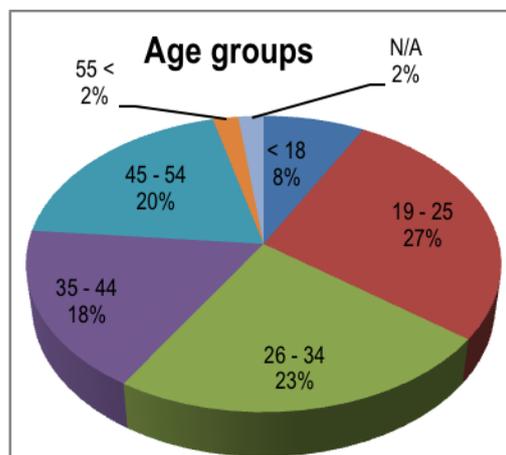


Figure 1 Age of participants

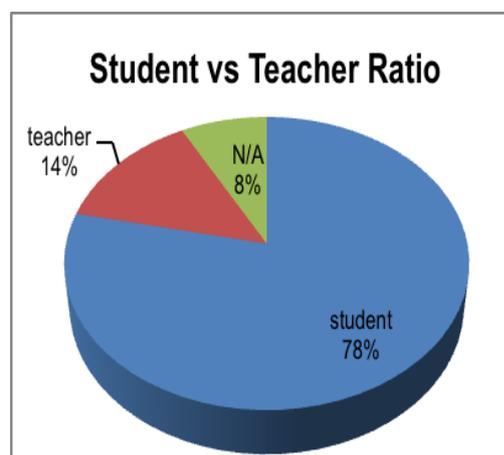


Figure 2 Occupation of participants

The data was analysed in three different ways. Method 1 was based on the Chi-squared test statistical significance. Observations with statistically significant number of ticks across all four dimensions (sig. < 0.05) were identified as having at least one preference. The strength of the preference was evaluated using the following criteria: Adjusted residual for a particular cell is in the range of (1.64 - 1.95) indicated a mild preference for that modality, adjusted residual for a particular cell is in the range of (1.96 - 2.56) indicated a strong preference and adjusted residual for a particular cell of 2.57 and above

indicated a very strong preference for that modality. This approach for identifying the strength of a modal preference is identical to the one adopted by Fleming [8] although the standard residuals in the original research were derived using a t-test. Observations that did not have statistically significant dimensions across the 4 modes (i.e. sig >=0.05) were classified as using all 4 modes (VARK).

Method 2 is similar to method one but rather than look at the relative strength across all 4 dimensions as measured by significance, it examined the adjusted residuals in each cell for significance across the positive proportion of the preferences. In this case all observations that had an overall significance level less than 0.05 as well as observations that had at least one statistically significant residual cell were classified as having strong preference. However, if a particular cell contained significantly fewer than expected number of ticks, this cell was ignored (i.e. it was not taken to indicate a lack of preference in that modality).

Method 3 is adopted the same approach as Method 2 but considered significant adjusted residuals both in the positive and in the negative sense so that having significantly fewer ticks on a dimension was interpreted as lack of preference in that modality and therefore the preference was removed.

The result from the three methods of quasi-ipsative scoring were compared with standard VARK method of scoring and it can be seen that the degree of agreement between the new and old scoring approaches increased as the information taken into account on making the recommendation increased. (Table 2).

Table 1 Overall agreement of proposed scoring methods with original scoring approach

Row Labels	Method 1		Method 2		Method 3	
match	30	59%	35	69%	39	76%
not match	21	41%	16	31%	12	24%
Grand Total	51	100%	51	100%	51	100%

The discriminatory power across modalities and sensitivity of the different scoring techniques also increased as the factors taken into considerations increased. Method 1 was very poor at predicting bimodal and tri-modal preferences although it is relatively accurate at predicting users with unimodal preferences (77%). However, the method overestimated significantly the proportion of users with multimodal preferences. Method 2 does less well in predicting users with unimodal preferences but is more accurate in identifying multimodal users. Neither Method 1, nor Method 2 identify any tri-modal preferences. Method 3, which considers both positive and negative preferences, performs better than Method 1 and Method 2 in identifying users with both unimodal and multimodal preferences. Furthermore, Method 3 identified a number of users with tri-modal preferences, although the categories identified as significant are different from the categories identified using the original t-test method.

Table 2 Agreement of proposed scoring methods with original scoring approach by number of learning preferences

Row Labels	Original Method	Method 1	Method 2	Method 3
1-preference	22	13 (77% match)	30 (67% match)	23 (83% match)
2-preferences	3	1 (100% match)	1 (100% match)	1 (100% match)
3-preferences	7	0 (0% match)	0 (0% match)	13 (54% match)
4-preferences	19	37(52% match)	20 (70% match)	14 (93% match)
Grand Total	51	51	51	51

3. Discussion

The application of the quasi-ipsative approach using a Chi-squared test to identify the relative strength of input mode preferences for Visual, Auditory, Read/Write and Kinaesthetic information lead to different classification of preferences from the original t-test approach. The quasi-ipsative approach was less sensitive in identifying bi-modal and tri-modal preferences and its recommendations varied significantly from the recommendation made using a t-test. This was due to the fact that very few user profiles were identified as having statistically significant differences between their preferences overall. The Chi-squared test was less likely to flag statistically significant differences compared to the t-test. This is not surprising since the chi-squared test is designed for non-parametric data and relaxing the assumption of that the number of ticks are normality distributed throughout the population leads to an increase in the expected standard error term. Furthermore, as mentioned above the t – test assumes that the data gathered is measured on at least approximately interval scale and counted data does not necessarily satisfy this assumption.

The instrument's sensitivity and discriminatory power could be improved significantly by considering individual cell significance and identifying individual cells with significantly more and/or fewer than the expected number of ticks (Methods 2 and 3). However, adopting Methods 2 and 3 violates the statistical rigor of the Chi-Squared approach as a cell in a table could have more/fewer than expected counts simply by chance. On the other hand, this approach replicates very closely the t-test approach taken in the original scoring of the questionnaire which also does not require a significant results across the categories but instead uses 1, 2 and 3 standard deviations from the mean as the indicator for the preference strength. Given that the data is nominal in nature, the Chi-squared test is a better alternative for data analysis as it makes fewer assumptions about the properties of the data.

The adoption of a quasi – ipsative approach also deals with potential classification discontinuities brought on by the use of averages that are not monotonically increasing. The original and new classification the individuals highlighted in Table 1 are displayed in Table 4 below. The new classifications handle some of the inconsistent classifications but again highlights the issue with the sensitivity of the alternative scoring approach as for example individual E is identified as having 4 modalities of equal strength, while the original method identified that user with a tri-modal preference.

Table 3 Original vs. New scoring approach comparison

	V	A	R	K	Total ticks	Original scoring	New Scoring (Method 3)
Individual A	0	3	8	4	15	strong R	strong R
Individual B	0	3	8	5	16	mild R	strong R
Individual C	2	4	5	2	13	AR	VAR K
Individual D	2	4	5	3	14	VAR K	VAR K
Individual E	2	5	4	6	17	AR K	VAR K
Individual F	2	5	4	7	18	mild K	VAR K

The introduction of quasi – ipsative scoring may have led to exaggeration of the negative preferences of individuals. By definition, the new classification is fixed sum, i.e. adding a tick to one category means that a tick was not given to another. The benefit from this approach is that it allows interpretation to both positive and negative preferences. So if a person had not ticked a V we can conclude that they are not visual in that scenario. However, the original test was not designed to identify negative preferences and it does not contain any control questions for negative preferences. Therefore, despite its ability to provide additional information, adopting the quasi-ipsative approach and using it to identify the lack of preferences without undergoing validation is not recommended.

4. Conclusions

This paper demonstrated the application of a quasi-ipsative scoring approach to assess the relative strengths of individual preferences in a VARK style questionnaire. The results suggest that the quasi-ipsative chi-squared based approach does not appear to be as sensitive as the original t-test approach in identifying significant modalities. In order to increase the sensitivity of the test, the requirement for overall test significance has to be relaxed and individual cells considered in the analysis. This potentially limits the validity of the proposed approach and as a cell could have a larger than expected number of observations purely by chance. The findings also put some doubt on the statistical validity of the original t-test approach as it also uses the deviation from the means (in the form of standard deviation) rather than statistical significance, as a tool for assessing the strength of individual preferences. The causes of the discrepancies between the two scorings techniques need to be examined further, before recommendations can be made on which approach is better suited to identifying the strength of individual preference for a information input modality.

References

- [1] N. D. Fleming and C. Mills, "Not another inventory, rather a catalyst for reflection," 1992.
- [2] K. R. Scherer, "What are emotions? And how can they be measured?," *Soc Sci Inform*, vol. 44, no. 4, pp. 695–729, Dec. 2005.
- [3] W. L. Leite, M. Svinicki, and Y. Shi, "Attempted Validation of the Scores of the VARK: Learning Styles Inventory With Multitrait-Multimethod Confirmatory Factor Analysis Models," *Educ Psychol Meas*, vol. 70, no. 2, pp. 323–339, Mar. 2010.
- [4] J. A. Krosnick, "Response strategies for coping with the cognitive demands of attitude measures in surveys," *Appl Cognitive Psych*, vol. 5, no. 3, pp. 213–236, May 1991.
- [5] J. M. Batista-Foguet, B. Ferrer-Rosell, R. Serlavós, G. Coenders, and R. E. Boyatzis, "An Alternative Approach to Analyze Ipsative Data. Revisiting Experiential Learning Theory," *Front. Psychol.*, vol. 6, pp. 575–10, Nov. 2015.
- [6] A. Brown and A. Maydeu-Olivares, "Item Response Modeling of Forced-Choice Questionnaires," *Educ Psychol Meas*, vol. 71, no. 3, pp. 460–502, May 2011.
- [7] J. F. Salgado, N. Anderson, and G. Tauriz, "The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis," *J Occup Organ Psychol*, vol. 88, no. 4, pp. 797–834, Dec. 2014.
- [8] N. D. Fleming, "I'm different; not dumb Modes of presentation (V.A.R.K.) in the tertiary classroom," presented at the Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia, 1995, vol. 18, pp. 308–313.