US-China Education Review A, August 2018, Vol. 8, No. 8, 332-340
doi: 10.17265/2161-623X/2018.08.002

Reliability of the VARK Questionnaire in Chinese Nursing Undergraduates

Julia S. W. Wong
Tung Wah College, Kowloon, Hong Kong
Kenny C. W. Chin
Stat Solutions Co., Kowloon, Hong Kong

All humans learn about the world through five senses as they are the first contact point when information enters the brain. The VARK learning style inventory is the sole assessment tool to measure an individual's learning preference in terms of senses. VARK represents visual (V), aural (A), read/write (R), and kinaesthetic (K). The aim of this study was to examine the reliability of VARK questionnaire version 7.8 in Chinese learners with four multitrait-multimethod confirmatory factor analysis (MTMM-CFA) models for further research study. A total of 177 Chinese nursing undergraduates were recruited. The results showed that correlated trait-correlated uniqueness (CTCU) was the best-fit model to the VARK scores and the reliability estimates for the scores of the VARK subscales ranged from 0.69-0.84. Among these 177 nursing students, there were 46 (26%) visual learners, 27 (15.3%) aural learners, 11 (6.2%) read/write learners, 21 (11.9%) kinaesthetic learners, and 72 (40.7%) multi-modal learners.

Keywords: VARK, learning styles, learning preference, confirmatory factor analysis, multitrait-multimethod

Introduction

Human beings have five senses, i.e., vision, touch, taste, hearing, and smell, and we learn about and understand our world through these five senses since childhood. Several renowned scholars have already linked human senses to learning for decades. For instance, the Atkinson and Shiffrin's model (1968) illustrated that our senses are the first contact point, e.g., light, sound, smell, heat, cold, and so forth, which brings information to the brain. Neil Fleming initially devised an inventory named the VARK questionnaire in 1987 which is based on the concept of his VARK model (Fleming & Mills, 1992). The latest revised version 7.8 was introduced in 2014. VARK represents visual (V), aural (A), read/write (R), and kinaesthetic (K), but includes five modalities which are visual, aural, read/write, kinaesthetic, and multimodal. This questionnaire is composed of 16 real-life scenario-based questions with four response options. It aims to help students, teachers, employees, customers, suppliers, and others to identify one's own learning style in terms of senses. Although there are controversies regarding over 70 different classification of learning styles (Coffield, Moseley, Hall, & Ecclestone, 2004), senses are innate routes of human learning in daily life and this proposition is highly accepted by scholars. Hence, the VARK questionnaire is one of the more popular inventories that researchers

Julia S. W. Wong, Ed.D., MN, BH.Sc., FHKAN, RN, School of Nursing, Tung Wah College.
Kenny C. W. Chin, CStat (RSS), M.Sc., B.Sc., B.A., Stat Solutions Co.

commonly use in the educational context. Leite, Svinicki, and Shi (2010) proved that the reliability of the VARK questionnaire for Neil Fleming with 15,136 students from the United States. However, scant empirical work has been implemented in the Chinese population. Given Leite et al. (2010) asserted that "Cronbach Alpha would underestimate the reliability of the VARK scores" (p. 33), this study followed their work by using four multitrait-multimethod confirmatory factor analysis (MTMM-CFA) models.

The Aim of the Study

The researcher of this study intended to conduct a research study, which is about the effects of a teaching method among students in different sensory modality groups, i.e., visual, aural, read, kinaesthetic, and multi-modal. Hence, the VARK questionnaire was used to categorize a group of Chinese nursing undergraduates enrolled in the same discipline course before moving on to the implementation of the abovementioned study. Thereby, the aim of the current study was to measure the reliability of this tool in the Chinese population.

Method

The 16-item VARK questionnaire version 7.8 was distributed to a total of 200 second and third years nursing undergraduates in the first lecture of a medical-surgical nursing course in 2016. All of them were local Chinese in Hong Kong. Given that English is the chief medium of instruction in higher education in Hong Kong, all undergraduates are required to attain a certain level of English in a public examination. Thereby, they should be capable of comprehending the VARK questionnaire.

Given that the respondents could select more than one option for each of the questions and each option represents one learning style preference, i.e., visual (V), aural (A), read/write (R), and kinaesthetic (K), VARK can be viewed as a questionnaire comprises 16 testlets with four dichotomous items each. Therefore, the correlations among items within testlets are a type of method effect. As a result, MTMM-CFA is used to model the method effects. Multitrait-multimethod (MTMM) is a design in which multiple traits are measured by multiple methods, where the original framework comes from the work of Campbell and Fiske (1959). MTMM analysis based on Confirmatory Factor Analysis (CFA) was based on the early work of Widaman (1985). The model included latent factors for both traits and methods, where correlation between traits and methods was not allowed. Leite et al. (2010) helped Fleming to examine the dimensionality of our VARK by using the four MTMM-CFA models: correlated trait-correlated method (CTCM), correlated trait-correlated uniqueness (CTCU), correlated-trait-uncorrelated method (CTUM), and correlated trait-correlated methods minus one (CT-C(M-1)), because they asserted that "Chronbach Alpha would underestimate the reliability of the VARK scores" (p. 33). Therefore, this study followed the work of Leite et al. and tested our data with four models. The analysis was conducted by Mplus Version 7.

Since the items of the VARK scale are dichotomous, the mean- and variance- adjusted weighted least squares estimator (WLSMV) was used. For the assessment of model fit, the researchers use numerous goodness-of-fit indicators to assess a model. Some common fit indexes include Chi-Square Test, Comparative Fit Index (CFI), Non-Normed Fit Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). However, Chi-Square Test as a measure of fit in a structural equation model is not recommended due to its sensitivity to sample size (Hair, Anderson, Tatham, & Black, 1998; Tabachnick & Fidell, 2007). Instead, the relative Chi-Square which equals the Chi-Square index divided by the degrees of freedom is recommended. This index might be less sensitive to sample size. The criterion for acceptance varies across researchers, ranging from less than 2 (Ullman, 2001) to less than 5 (Schumacker & Lomax, 2004).

Regarding the criteria of the Comparative Fit Index (CFI) and Non-Normed Fit Index (TLI), a value of at least 0.90 is required to accept a model, while a value of at least 0.95 is required to judge the model fit as "good" (Holmes-Smith, Coote, & Cunningham, 2004). Another approach to test the model fit is to accept a model that approximates the true model through the RMSEA index. RMSEA of less than 0.05 indicates a close fit, and values between 0.05 and 0.08 indicate an acceptable fit, while SRMR with a value below 0.08 is indicative of acceptable fit. As the models cover four traits and 16 testlets, agreement on the fit from many fit indices is difficult to obtain with large multifactor models (Marsh, Hau, & Grayson, 2005). All statistical tests were two-tailed and variables were considered significant at a level of 0.05.

Results

A sample of 177 (88.5%) students completed questionnaires were received (male N = 51 and female N = 126) for analysis. Among these 177 students, 46 (26%) were visual learners, 27 (15.3%) were aural learners, 11 (6.2%) were read/write learners, 21 (11.9%) were kinaesthetic learners, and 72 (40.7%) were multi-modal learners. There were no missing data. The percentage of the responses on the four items for each of the questions is shown in Table 1. Tetrachoric correlation matrix would be used for testing with the following four MMTM-CFA hypothesized models.

Table 1

Percentage of Items Responded Among Questions (N = 177)

	Item 1		Item 2		Item 3		Item 4	
	N	%	*N*	%	*N*	%	*N*	%
Q1	51	28.8	143	80.8	29	16.4	31	17.5
Q2	97	54.8	53	29.9	60	33.9	115	65.0
Q3	32	18.1	86	48.6	53	29.9	101	57.1
Q4	137	77.4	63	35.6	53	29.9	29	16.4
Q5	70	39.5	70	39.5	52	29.4	71	40.1
Q6	76	42.9	31	17.5	113	63.8	95	53.7
Q7	27	15.3	87	49.2	44	24.9	140	79.1
Q8	65	36.7	107	60.5	51	28.8	58	32.8
Q9	66	37.3	102	57.6	57	32.2	49	27.7
Q10	95	53.7	64	36.2	59	33.3	67	37.9
Q11	40	22.6	71	40.1	70	39.5	74	41.8
Q12	86	48.6	42	23.7	72	40.7	74	41.8
Q13	81	45.8	50	28.2	81	45.8	132	74.6
Q14	39	22.0	91	51.4	67	37.9	84	47.5
Q15	105	59.3	64	36.2	76	42.9	88	49.7
Q16	87	49.2	66	37.3	79	44.6	71	40.1

Note. * Respondents can select more than one item for each of questions.

Model 1: Correlated Trait-Correlated Method (CTCM)

The model (see Figure 1) includes both traits and methods factors, and allows for correlations among traits and among methods. The initial model cannot be generated as a non-positive definite latent variables covariance matrix was found. The source of the problem was a perfect linear dependency between some latent variables. Therefore, the CTCM model should be modified. In the modified model, one factor loading of some method factors were fixed to 1 and error correlation between some method factors were fixed to zero, in addition to the factor variances which were fixed to 1 to set the scale of the latent variables. This modified CTCM model was achieved with reasonable modification indices and the fitness indices were λ^2 (1,840) = 2,058.436, p < 0.001, relative Chi-Square (λ^2/df) = 1.12, CFI = 0.687, NFI = 0.657, RMSEA = 0.026, and SRMR = 0.129.

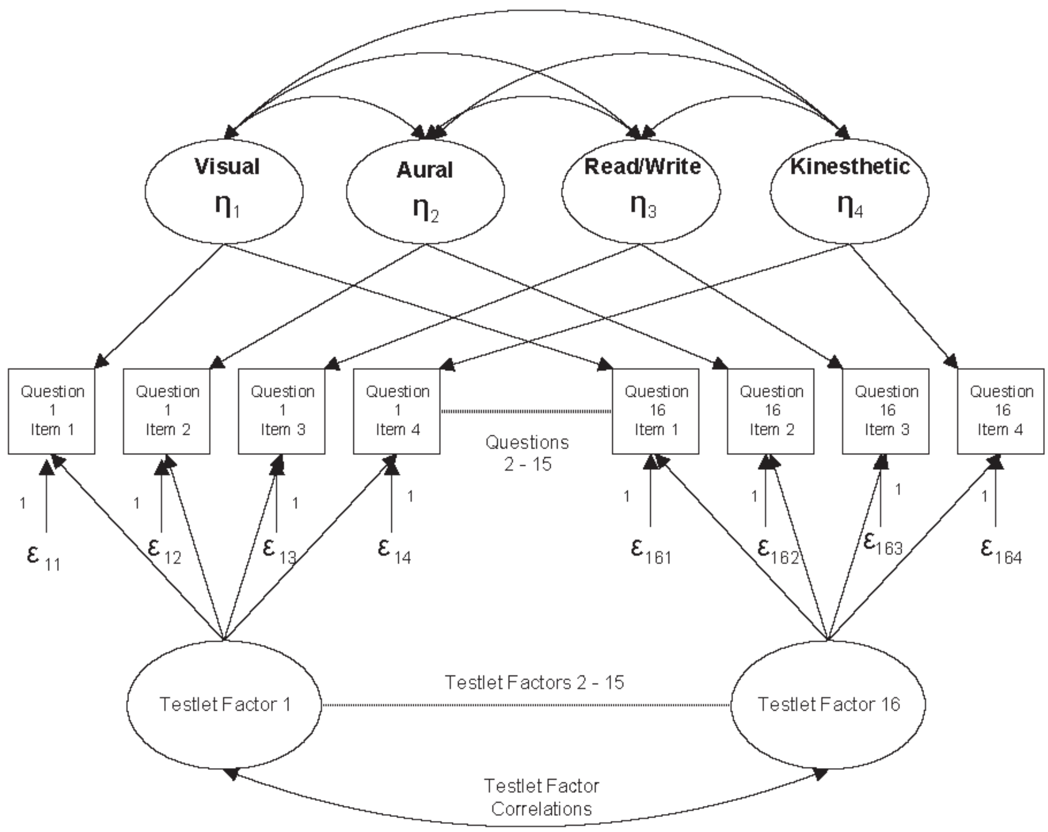

Figure 1. Model 1: Correlated trait-correlated method (CTCM).

Model 2: Correlated Trait-Correlated Uniqueness (CTCU)

CTCU (see Figure 2) was proposed by Marsh et al. (2005), in which no latent factor for methods was specified. This implies that the method effects are confounded with the correlated errors. Furthermore, correlations between methods are not allowed. However, the correlated residuals for each set of observed variables measuring the same method of measurement are considered to reflect these method effects. The initial model cannot be generated as a non-positive definite residual covariance matrix was found. The source of the problem was a perfect linear dependency between some errors of items of the kinaesthetic scale. Hence, the model was revised by fixing the error correlations to solve the problem. Among 96 error correlations of items within testlets, 29 were fixed. The revised model was achieved with reasonable modification indices and the fitness indices were λ^2 (1,883) = 2,021.625, p < 0.05, relative chi-square (λ^2/df) = 1.07, CFI = 0.802, NFI = 0.788, RMSEA = 0.020, and SRMR = 0.125.

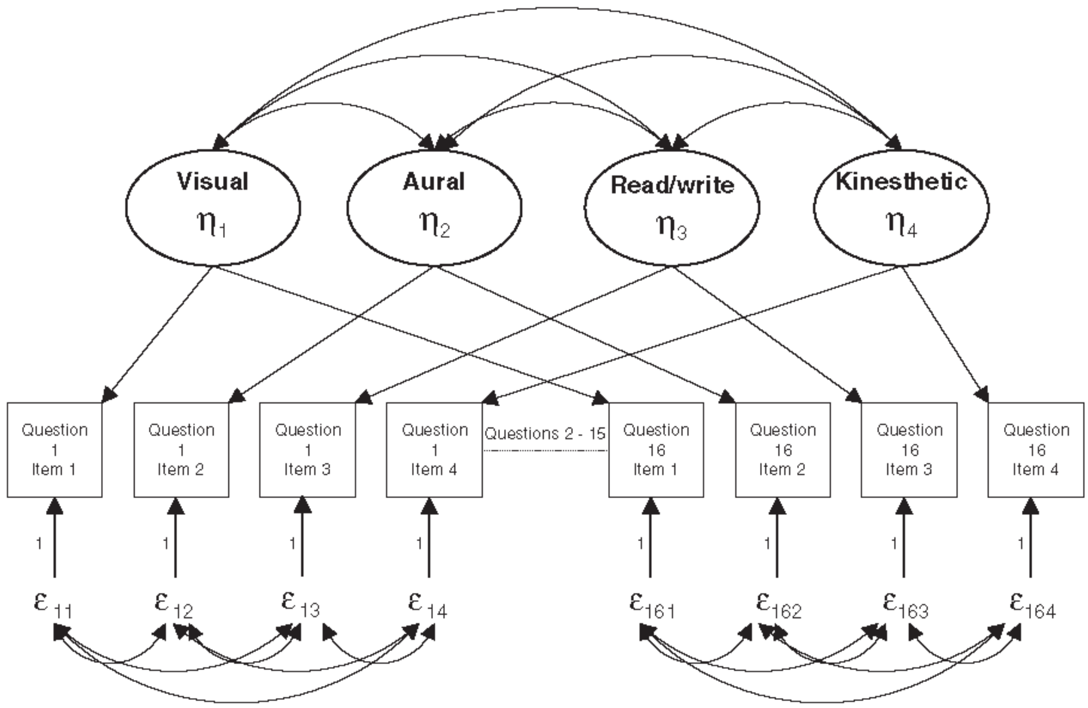

Figure 2. Model 2: Correlated trait-correlated uniqueness (CTCU).

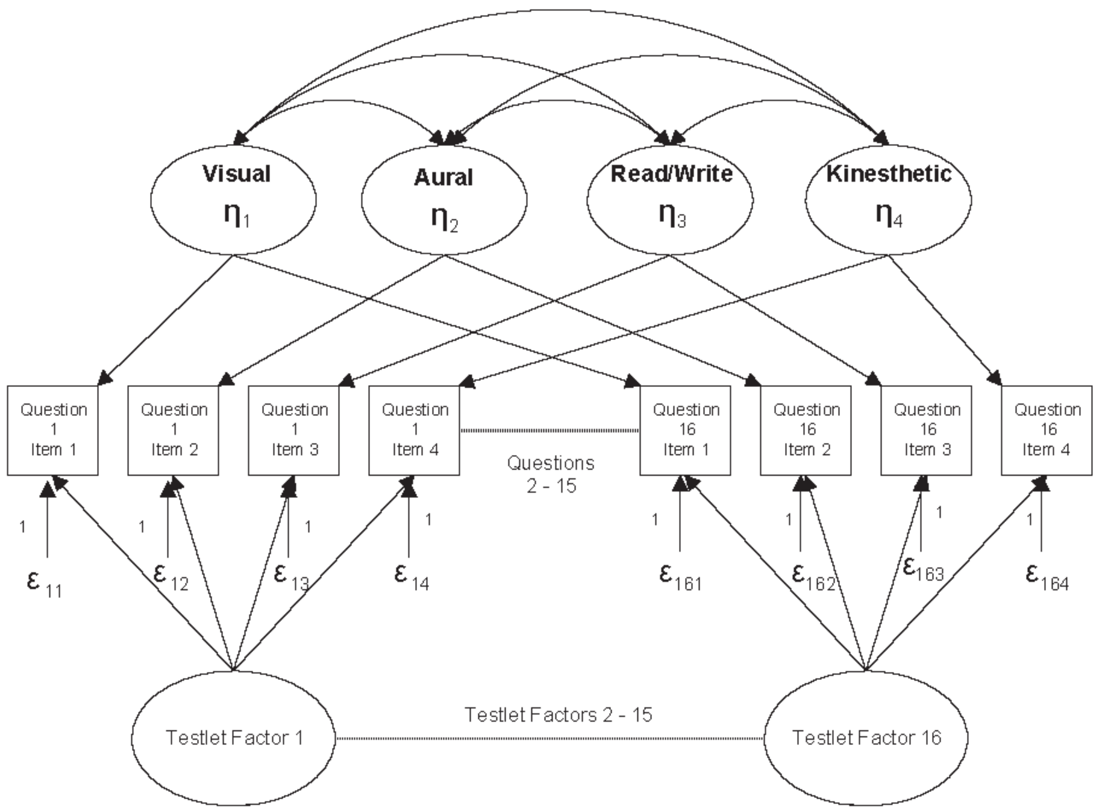

Figure 3. Model 3: Correlated trait-uncorrelated method (CTUM).

Model 3: Correlated Trait-Uncorrelated Method (CTUM)

CTUM (see Figure 3) is similar to CTCM, except that it did not have any specified correlations among the method factors. Goodness-of-fit results for this mode l were λ^2 (1,898) = 2,117.381, $p < 0.001$, relative Chi-Square (λ^2/df) = 1.12, CFI = 0.686, NFI = 0.666, RMSEA = 0.026, and SRMR = 0.133.

Model 4: Correlated Trait-Correlated Method Minus One (CT-C(M-1))

CT-C (M-1) (see Figure 4) was proposed by Eid (2000), in which the number of method factors is specified to be m-1, where m is the number of methods in the design. We ran the model one by one by changing the reference method every run. The results show that using the first question as a reference method gave the best goodness-of-fit indices with λ^2 (1,837) = 2,058.579, $p < 0.001$, relative Chi-Square (λ^2/df) = 1.12, CFI = 0.683, NFI = 0.652, RMSEA = 0.026, and SRMR = 0.129.

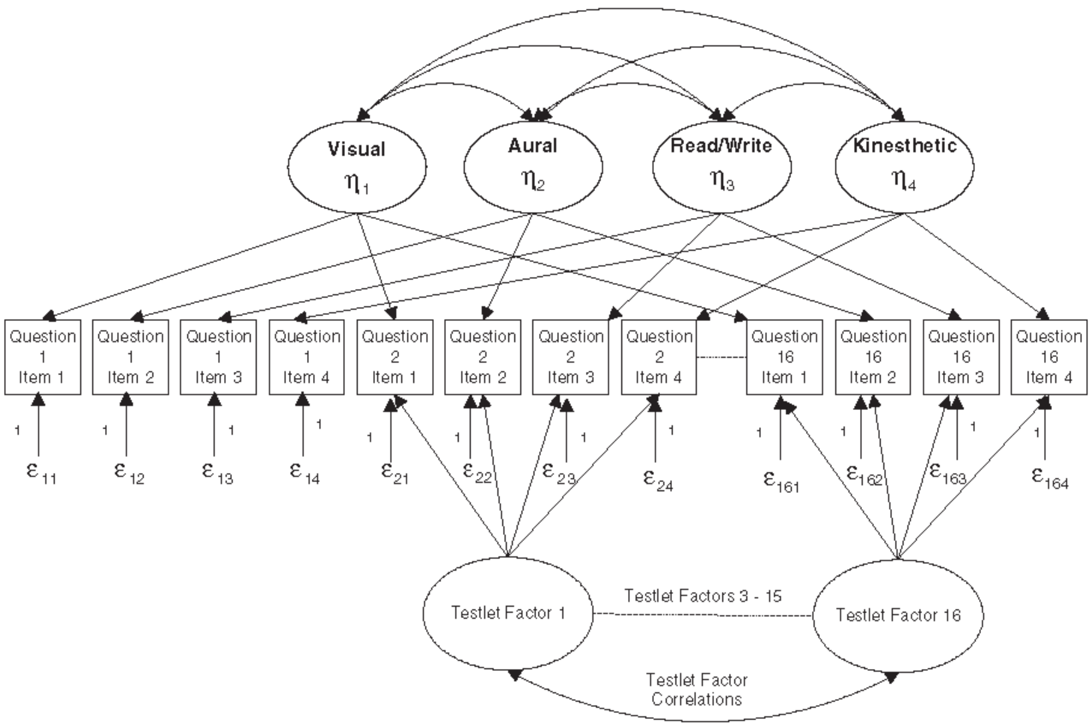

Figure 4. Model 4: Correlated trait-correlated method minus one (CT-C(M-1)).

A summary of model fit information for the above four MTMM-CFA models is shown in Table 2. It is found that the relative Chi-Square, RMSEA, and SRMR, but not TLI and CFI, support the fit of the four models. Among them, the CTCU model is the best fit to the data, which is consistent with the result reported from Leite et al. (2010). Furthermore, a CFA model with a diagonal residual covariance matrix was conducted to show the consequences of inappropriately modelling scores of testlets by failing to account for method effects. The goodness-of-index is λ^2 (1,837) = 2,259.625, $p < 0.001$, relative Chi-Square (λ^2/df) = 1.16, CFI = 0.551, NFI = 0.535, RMSEA = 0.030, and SRMR = 0.140, which is a poor fit model compared with CTCU. Since this model is constructed within the CTCU model, the resulting Chi-Square difference test is λ^2 (63) = 648.826, $p < 0.001$, which indicated that the CTCU model is statistically significant in the improvement in fit with the data when compared to a CFA model ignoring method effects.

The results of the CTCU model show that the standardized loadings of the items on the VARK factors ranged from 0.063 to 0.643, with mean loadings of 0.375, 0.432, 0.408, and 0.346 for the visual, aural, read/write, and kinaesthetic factors, respectively. Thus, the correlations between the VARK factors were moderate to strong in magnitude (see Table 3). The reliability estimates for the scores of the VARK subscales were 0.73, 0.79, 0.84, and 0.69 for visual, aural, read/write, and kinaesthetic subscales, respectively, which means the reliability of VARK questionnaire in this population group is acceptable.

Table 2

Summary of Goodness-of-Fitness Statistics for CFA-MTMM Models

Model	λ^2	Degree of Freedom (*df*)	*p*	λ^2/df	CFI	TLI	RMSEA	RMSEA 90% confidence interval (CI)	SRMR
1. CTCM	2,058.436	1,840	< 0.001	1.12	0.687	0.657	0.026	0.019 0.032	0.129
2. CTCU	2,021.625	1,883	< 0.05	1.07	0.802	0.788	0.020	0.010 0.027	0.125
3. CTUM	2,117.381	1,898	< 0.001	1.12	0.686	0.666	0.026	0.018 0.032	0.133
4. CT-C(M-1)	2,058.579	1,837	< 0.001	1.12	0.683	0.652	0.026	0.019 0.032	0.129

Table 3

Standardized Factor Loadings and Factor Correlations for the VARK[+]

Factor loadings	Visual	Aural	Read/Write	Kinaes thetic
Q1	0.104[b]	0.540	0.197[b]	0.074[b]
Q2	0.333	0.421	0.575	0.242
Q3	0.198[b]	0.369	0.513	0.254
Q4	0.439	0.392	0.353	0.432
Q5	0.249	0.441	0.270	0.004[b]
Q6	0.606	0.352	0.316	0.571
Q7	0.594	0.643	0.512	0.444
Q8	0.181[b]	0.590	0.503	0.063[b]
Q9	0.298	0.609	0.339	0.334
Q10	0.423	0.268	0.469	0.515
Q11	0.253[b]	0.483	0.170[b]	0.266
Q12	0.354	0.464	0.267	0.374
Q13	0.597	0.450	0.546	0.581
Q14	0.565	0.202	0.551	0.373
Q15	0.446	0.452	0.391	0.595
Q16	0.366	0.242	0.555	0.417
Factor correlations				
Visual	-	0.574	0.539	0.570
Aural	-	-	0.706	0.744
Read/write	-	-	-	0.416

Notes. [+] Correlated trait-correlated uniqueness (CTCU) model;

[b] Not statistically significant ($p < 0.05$).

Discussions and Conclusions

Given that there are limited studies exploring the reliability of VARK with MTMM-CFA, the results of this study can be used to compare with Leite et al. (2010) only. Both studies found that CTCU is the best-fit model to the VARK scores among the four MTMM-CFA models, even though this study had a relatively small sample size. In addition, acceptable reliability coefficients based on the CTCU model were obtained in both studies. When a CFA model with a diagonal residual covariance matrix was compared with the CTCU model, both studies showed that the CTCU model is better. In Leite et al. (2010), the reliability estimates for the scores of the VARK subscales were 0.85, 0.82, 0.84, and 0.77 for V, A, R, and K subscales with over 15,000 participants in United States. Although the reliability estimates of this study were a little bit lower than in Leite et al. (2010), i.e., 0.73, 0.79, 0.84, and 0.69 with 177 participants, they still remained within the acceptable range in which three of them were $\geq$ 0.7 and the rest was > 0.8. It may be concluded that the VARK questionnaire is also sensitive and reliable with a relatively small sample size of around 180 Chinese undergraduates. Further examination with a larger sample size is recommended.

With the help of VARK classification, the results revealed that the majority of the students were multimodal learners (41%) among the 177 students. This finding was the same as in other studies conducted in various healthcare professions, such as nursing (Alkhasawneh, Mrayyan, Docherty, Alashram, & Yiusef, 2008; Alkhasawneh, 2013), medicine (Nuzhat, Salem, Al Hamdan, & Ashour, 2013; Sinha, Bhardwaj, Singh, & Abas, 2013; Slater, Lujan, & DiCarlo, 2007; Urval et al., 2014), dentistry (Murphy, Gray, Straja, & Bogert, 2004), physiotherapy (Breckler Joun, & Ngo, 2009; Lujian & DiCarlo, 2006), and health science (Meehan-Andrews, 2009). It may indicate that many university students are capable of managing different situations or different kinds of knowledge with different sensory modalities in order to obtain the best outcome or it may relate to the nature of the discipline as the same phenomenon is only found in health-related or science subjects. This may not be the case in other disciplines.

The results of this study strengthened the use of VARK question version 7.8 in the Chinese population as there is limited study in this area. In addition, it allows researchers to extend the scope of research in the educational context. For instance, the researcher may use it to compare the differences between different sensory modality groups after adopting a new teaching method if she or he considers that it is a variable.

References

Alkhasawneh, E. (2013). Using VARK to assess changes in learning preferences of nursing students at a public university in Jordan: Implication for teaching. *Nurse Education Today, 3*3(22), 1546-1549.

Alkhasawneh, I. M., Mrayyan, M. T., Docherty, C., Alashram, S., & Yiusef, H. Y. (2008). Problem-based learning (PBL): Assessing students' learning preferences using VARK. *Nurse Education Today, 28*(5), 572-579.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic Press.

Breckler, J., Joun, D., & Ngo, H. (2009). Learning styles of physiology students interested in the health professions. *Advances in Physiology Education, 33*(1), 30-36.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminate validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning: A systematic and critical review*. London: Learning and Skills Research Centre.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*(2), 241-261.

Fleming, N. D., & Mills, C. (1992). Not another inventory, rather a catalyst for reflection. *To Improve the Academy, 11*, 137-155.

Gardner, H. E. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, US: Basic Cooks.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). New Jersey, NJ: Prentice-Hall.

Holmes-Smith, P., Coote, L., & Cunningham, E. (2004). Structural equation modeling: From the fundamentals to advanced topics. In *ACSPRI-Summer Training Program.* Canberra, Australia.

Leite, W. L., Svinicki, M., & Shi, Y. (2010). Attempted validation of the scores of the VARK: learning styles inventory with multitrait-multimethod confirmatory factor analysis models. *Educational and Psychological Measurement, 70*(2), 323-339.

Lujian, H. L., & DiCarlo, S. E. (2006). First-year medical students prefer multiple learning styles. *Advances in Physiology Education, 30*(1), 13-16.

Marsh, H. W., Hau, K., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares and J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275-340). Mahwah, NJ: Lawrence Erlbaum Associates.

Meehan-Andrews, T. A. (2009). Teaching mode efficiency and learning preferences of first year nursing students. *Nurse Education Today*, 29, 24-32.

Murphy, R. J., Gray, S. A., Straja, S. R., & Bogert, M. C. (2004). Student learning preferences and teaching implications. *Education Methodologies, 68*(1), 859-866.

Nuzhat, A., Salem, R. O., Al Hamdan, N., & Ashour, N. (2013). Gender differences in learning styles and academic performance of medical students in Saudi Arabia. *Medical Teacher, 35*(Suppl. 1), 78-82. doi:10.3109/0142159X.2013.765545

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Sinha, N. K., Bhardwaj, A., Singh, S., & Abas, A. L. (2013). Learning preferences of clinical students: A study in a Malaysian medical college. *International Journal of Medicine and Public Health, 3*(1), 60-63.

Slater, J. A., Lujan, H. L., & DiCarlo, S. E. (2007). Does gender influence learning style preferences of first-year medical students? *Advance in Physiology Education, 31*, 336-342.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York: Allyn and Bacon.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick and L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed., pp. 653-771). Needham Heights, MA: Allyn & Bacon.

Urval, R. P., Kamath, A., Ullal, S., Shenoy, A. K., Shenoy, N., & Adupa, L. A. (2014). Assessment of learning styles of undergraduate medical nursing students using a VARK questionnaire and the influence of sex and academic performance. *Advances in Physiology Education, 38*, 216-220.

VARK Copyright. (1998). Neil D. Fleming, Christchurch, New Zealand and Charles C. Bonwell. *Green Mountain Falls, Colorado, 80819*(719), 684-9261.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1-26.