Is context the hidden spanner in the works of educational measurement: exploring the impact of context on mode of learning preferences

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Is context the hidden spanner in the works of educational measurement: exploring the impact of context on mode of learning preferences

Elena Fitkov-Norris[1] and Ara Yeghiazarian

Kingston Business School, Kingston University, Kingston Hill, Kingston upon Thames KT2 7LB, UK

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

Abstract. This paper examines the VARK learning preferences evaluation tool for evidence of context effect as a result of the questionnaire design using a common scenario stimulus to elicit learning preferences. Confirmatory Factor Analysis (CFA) is used to test the overall model fit of a multidimensional hierarchical model which included a context scenario effect and a non-hierarchical model which does not explicitly model a context effect. The results show that half of the scenarios in the VARK learning preferences questionnaire have significant context effect, which in turn introduce response bias. The presence of context response bias in some of the questions offers a possible explanation to the lack of consistent empirical evidence for the existence of preference styles flagged by educators, which contradicts findings from neuroscience and psychology of a convergence effect between learning modality preferences and learning approaches. The findings have significant implications for designers of learning preference tools in particular and educational measurement tools in general as they highlight the risks of the inadvertent introduction of possible response bias when implementing scenario-based evaluation tools.

1. Introduction

Measurement in educational research is continuing to thrive. For example, standardised testing in core subjects such as reading and mathematical attainment is now well established and applied throughout the entire length of educational career. In the meantime, the pervasive use of educational technology is making more data readily available, potentially shedding light on best educational practices through the use of meta analytic studies. However, the increased sample sizes and large data sets have failed to produce the empirical evidence to support some well-established and long held beliefs in what are best practices in education. For example, meta analytic studies into learning styles have failed to find empirical support to the hypothesis that adjusting the mode of presentation materials to the preferred mode of learning preference, also known as convergence theory, will ultimately lead to better educational attainment [1]. Despite the lack of empirical evidence, this learning preferences myth is thriving [2] and articles validating tools for assessing the strength of learning preference for visual, auditory, read/write and kinesthetic (VARK) inputs continue to be published [3] possibly due to the intuitive appeal of these concepts.

---

[1]   To whom any correspondence should be addressed.

However, in direct contradiction to the empirical meta analytic findings, evidence has emerged from neuroscientific and psychological direct observational studies into information processing. For example, functional Magnetic Resonance Imaging (fMRI) scans have shown that the convergence theory holds true for people with strong visual preference, who have to convert words into pictures when presented with text input in order to understand the input. Similarly, the convergence theory holds for people with strong read/write text preference, who have to convert pictures into text to help with comprehension [4]. Further support to the existence of a convergence effect in preference for visual or text based (read/write) information intake is provided by eye-tracking studies in psychology, showing that visualizers (students with a preference for pictures and diagrams) and verbalisers (students with preference for text) generally examine the areas on the screen where information is presented in line with their preference and for instance visualizers tend to focus for longer on information rich areas that hold visual cues (such as diagrams) than verbalisers [5]. Conversely, verbalisers tend to focus on areas that hold textual cues (such a text). The presence of this contradictory evidence suggests that there could exist a confounding variable(s) or process(es) that is unaccounted for in the design of the studies.

One possible explanation for the apparent contradiction between empirical evidence gathered in the classroom and functional and processing empirical evidence gathered by psychologists and neuroscientists in controlled lab conditions is the impact of context. The confounding effect of context has been reported in a number of meta analytic studies in a number of diverse areas, such as energy efficiency [6], healthcare [7], ecology [8] and advertising [9], leading to apparently contradictory effects of interventions and affecting the significance of hypothesized relationships.

This article tests whether context could play a significant confounding role in the measurement of learning preferences using questionnaires and thus potentially leading to contradictory findings to the ones made by direct observation, this article examines the responses given to the VARK learning preferences questionnaire for consistency across 16 different scenario contexts with a view of providing some insights into the impact of context on the strength of self-reported mode of learning preference. Confirmatory Factor Analysis (CFA) is used to compare the fit of a model that takes into account the context effect and a model that do not explicitly take into account the context effect.

The article is organised as follows: the next section gives a formal definition of context and an overview of the literature on the empirical evidence for the impact of context and other design factors that affect the reliability self-reported measuring instruments in the area of educational research. Section 3 discusses the methodology used to evaluate the context effect for the VARK preference measurement tool. Section 4 reports the results from the analysis and discusses the implications from the findings in the context of educational measurement. The final section summarises the research findings and makes recommendations for further research.

2. Context in educational measurement

2.1. Definition of context

Awareness of the role of context in education is growing and educators have recommended approaches such as Problem Based Learning (PBL) in a range of contexts as a better way to encourage acquisition and retention of transferable knowledge [10, 11]. However, despite this growing awareness, there is no universally accepted definition of context in educational research and researchers utilise a variety of terms such as context effect in decision sciences when evaluating different alternatives [12] or context specificity to refer to the impact of context on educational attainment [11, 13] or personality tests [14]. For example, context specificity has been demonstrated across a range of teaching related situations such as perceived teacher enthusiasm [13], learning motivation [15], doctoral supervision [16]. More recent works such as [16, 17] has highlighted the issues arising from decontextualization and suggests that applying the lens of decontextualization to educational research leads to poor understanding of the complexities of the educational interactions. In fact, empirical evidence is emerging showing that context is best viewed as a complex multidimensional construct, rather than a single factor [13, 18]. To reflect the multidimensional nature of context this this paper will follow the definition by Parker et. al.

[19, page 46]. who define context as something that "… can be derived from anything that is significant in a given moment and potentially including the environment, an item within that environment, a user, or even an observer." This is an all-encompassing definition that can cover a range of potential situations and scenarios, as well as reflect various research and measurement tool designs.

2.2. The VARK learning preference measurement tool

The VARK learning modalities questionnaire measures the degree to which learners prefer to absorb information in visual (V), auditory (A), Read/Write (R) and Kinaesthetic (K) form in different educational settings [20]. Visual form includes preference for diagrams, charts to explain ideas as opposed to text which is the preferred method for people with high Read/Write preference or spoken words which are the preference for people with high Auditory preference. People with high kinaesthetic preference prefer to use experience, examples or trial and error. The VARK questionnaire (http://vark-learn.com/the-vark-questionnaire/) consists of 16 items and each item has four possible answer, each indicating a particular preference. Users are encouraged tick all items that apply in each case. Two questions taken from an earlier version (2014) of the questionnaire have been shown below:

Q1: A group of tourists wants to learn about the parks or wildlife reserves in your area. You would:
a. show them maps and internet pictures. (answer represents V)
b. talk about, or arrange a talk for them about parks or wildlife reserves. (answer represents A)
c. give them a book or pamphlets about the parks or wildlife reserves. (answer represents R)
d. take them to a park or wildlife reserve and walk with them. (answer represents K)

Q2: You have a problem with your heart. You would prefer that the doctor:
a. used a plastic model to show what was wrong. (answer represents K)
b. showed you a diagram of what was wrong. (answer represents V)
c. gave you something to read to explain what was wrong. (answer represents R)
d. described what was wrong. (answer represents A)

Following on from the definitions of context provided in section 2.1, each question is positioned in a particular context (encountering a group of tourists (Q1) or having a problem with one's heart (Q2)) that could potentially bias the responses of participants. For, example, in the context of encountering a group of tourists, an introvert may choose to act differently to an extrovert, despite their actual learning modality preference. In the context of concern about one's heart, a respondent may wish to utilise all four options given, in order to maximise their understanding. Context is also present in the range of response options. For example, a male or a female respondent could have different risk propensity to walking with strangers, which could affect their response to item Q1(d). In fact, attempts to validate the questionnaire have shown that fitting a model that takes into account the scenarios leads to a better overall model fit [21], although the relative contribution of each scenario was not examined or reported.

Despite the potential to introduce bias, contextualization can have a positive impact on measurement and, for example, it was found to reduce the error variances in personality measurement instruments [14]. Research has also shown that the common stimuli (context) design across different dimensions can mitigate the negative impact of local item dependence although the single stimuli (context) design across a single dimension can introduce a significant bias [22]. Thus, untangling the impact of context on the validity of measurement tools is not trivial as the questionnaire design can also play a significant role in biasing responses.

3. Methodology

This research utilizes Fleming's VARK questionnaire [23] to indicate the mix and strength of students' preferred learning modes[2].

---

[2] This research used the 2014 version of the VARK Questionnaire. The Questionnaire has had a number of options and questions replaced as well as the wording of some options and question-stems changed and is now in version 8.0. Those changes will have affected some of the results shown.

The sample consists of responses by 17,413 participants who complete the VARK questionnaire on the VARK website between June and December 2014.

The majority of respondents were students (89 %) and female (65 %). 52 % of respondents were aged between 19 34 and the majority were classified as having a single preference mode (37 %), followed by respondents who utilized all 4 preferences (35 %). A smaller proportion of respondents utilized either 2 (15 %) or 3 modes (13 %) respectively.

The data were analysed using R version 3.4.4 (2018-03-15) -- "Someone to Lean On" [24] using the "mirt" package [25]. A single layer multidimensional confirmatory factor model which hypothesized the existence of the four main factors (Visual, Auditory, Read/Write and Kinesthetic) but did not take into account the existence of scenarios, was fitted to the data first. This was followed by the fitting of a hierarchical multidimensional model which considered the existence of the four main factors (Visual, Auditory, Read/Write and Kinesthetic) and 16 secondary scenario factors that introduced different contexts to each of the main dimensions. The strength of context effect for each scenario was evaluated using each secondary factor's eigen value. Figure 1 is a graphical representation of the examined two-tier hierarchical model with four primary latent VARK dimensions and 16 secondary or specific latent scenario dimensions, together with the 64-indicator variables (16 per each dimension). The non-hierarchical model consists of just four primary latent VARK dimensions, each with 16 indicator variables.
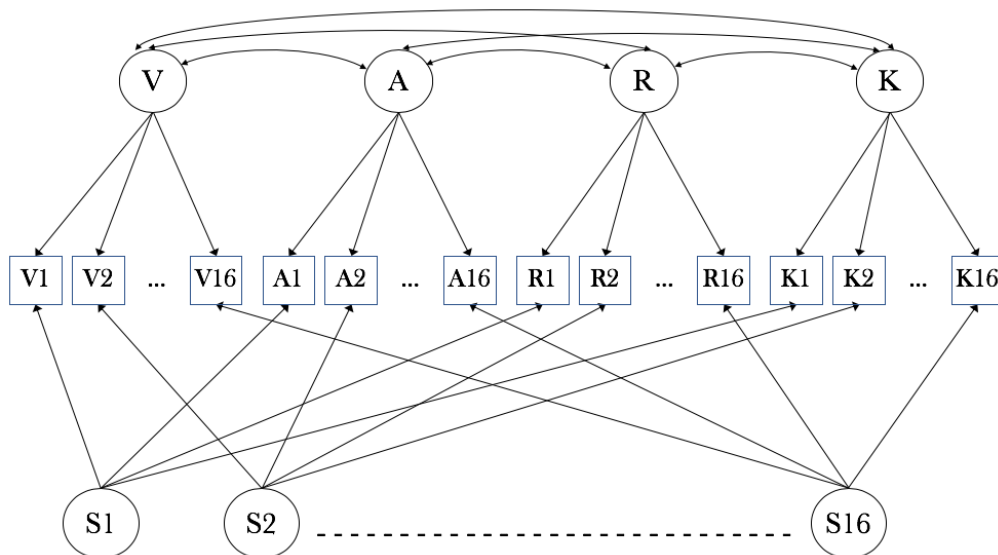

Figure 1. Two-tier hierarchical model with secondary context effects.

Following from [26] the conditional probability of correct endorsement response for a dichotomously scored item y_{ij} in a hierarchical model with p primary dimensions and s specific or secondary dimensions is given by:

$$P_j(y_{ij} = 1|\boldsymbol{\eta_i}, \xi_{is}, \boldsymbol{\theta}) = (1 + \exp\{-[\alpha_i(\boldsymbol{\theta}) + [\boldsymbol{\beta_i}(\boldsymbol{\theta})]'\boldsymbol{\eta_i} + \beta_{is}(\boldsymbol{\theta})\xi_{is}]\})^{-1} \tag{1}$$

where α_i is the intercept, $\boldsymbol{\beta_i}$ is the $p \times 1$ vector of item slopes on the primary factor vector $\boldsymbol{\eta_i}$, β_{is} is the item slope on specific factor ξ_{is} and $\boldsymbol{\theta}$ is a vector of all estimable and/or structural parameters in the two-tier model. The conditional probability for the incorrect/non-endorsement response is $P_j(y_{ij} = 0|\boldsymbol{\eta_i}, \xi_{is}, \boldsymbol{\theta}) = 1 - P_j(y_{ij} = 1|\boldsymbol{\eta_i}, \xi_{is}, \boldsymbol{\theta})$.

4. Results and discussion

The results from the two models are summarized in table 1. The two-tier hierarchical model with four primary dimensions for Visual, Auditory, Read/Write and Kinesthetics preferences and 16 secondary

context dimensions representing a different context fits the data better than the non-hierarchical multidimensional model that does not consider context effects. In particular, the two-tier hierarchical model has lower Root Mean Square Error of Approximation (RMSEA) with a 95% CI [.04159, .04217] meets the standard requirement of being below 0.7 [27]. Furthermore, the two-tier hierarchical model has a higher Tucker Lewis Index (TLI) which at .87366 is closer to the desired value of 1. In addition, the two-tier model also has a higher Comparative Fit Index (CFI) measure and significantly lower Bayesian Information Criterion (BIC) value. Both models have eigenvalues greater than 1 for the four main factors of Visual, Auditory, Read/Write and Kinaesthetic preferences, suggesting that the 64 items measure a 4-dimensional structure. The magnitude of the factor eigenvalues for the secondary factors in the hierarchical model is greater than 1 for eight out of the 16 scenarios, suggesting a significant context effect. Further textual analysis of the scenarios with significant context effect may provide valuable insights into the type of context that is prone to introduce bias and so inform good design practices, although it falls beyond the scope of this work.

Table 1. Model results comparison.

Model Fit Measures	Non-Hierarchical Model	Two-tier hierarchical model with context effects
RMSEA	0.05345	0.04188
RMSEA LCL	0.05316	0.04159
RMSEA UCL	0.05373	0.04217
TLI	0.77963	0.86467
CFI	0.78728	0.87366
AIC	1337249	1304551
BIC	1338290	1306089 [a]

[a] Significant at 1% level

The results show that the context effect is significant and present in half of the scenarios used in the VARK questionnaire. This suggests that the strength of the context effect on the expression of learning preferences is not ubiquitous but specific to each situation. The result is surprising on the one hand, given that prior research has indicated that utilising a single stimulus design across dimensions reduces inter-item dependence significantly [22]. Thus, by adopting a single stimulus design and measuring the strength of all four preferences (Visual, Auditory, Read/Write and Kinaesthetic) within each scenario context, should have mitigated the impact of context. However, on the other hand this finding is in line with empirical research in the field of education which has identified differential context effect functioning in a range of measurements such as teacher enthusiasm [13], student motivation in online distance learning courses [15], dissertation supervisory practices [16] and memory recall tasks [11].

The presence of context response bias in some of the questions also offers a possible explanation to the lack of consistent empirical evidence for the existence of convergence effect between learning modality preferences and learning approaches flagged by educators [1]. Given that the context is not ubiquitous, its presence is challenging to identify and therefore its impact in meta analytical studies my overpower the learning modalities or learning preferences convergence effect. In turn this will provide some justification for contradictory findings from neuroscience [4] and psychology [5] on the presence of such convergence effect in the case of people with strong learning modality preference. This finding has overarching implications for the future design decisions of preference measurement instruments such as the VARK questionnaire, suggesting that utilization of single stimulus design across multiple dimensions may be necessary, but not sufficient requirement to avoid the impact of context bias on user responses. The results emphasise the inherent complexity of building reliable measuring tools in the

presence of significant context effects and highlight the need for further investigation into the impact of context in educational research.

5. Conclusions

This paper examined the VARK learning preferences evaluation tool for evidence of context effect as a result of the questionnaire design using a common scenario stimulus to elicit learning preferences. The results suggest that half of the scenarios have significant evidence of context effect, which in turn introduce response bias. The presence of context effect has implications for the future design decisions of preference measurement instruments such as the VARK questionnaire, suggesting that utilization of single stimulus design across multiple dimensions may be not be sufficient to avoid the impact of context bias on user responses. These findings have significant implications for designers of learning preference tools in particular and educational measurement tools in general as they highlight the risks of introducing a significant response bias when implementing scenario-based evaluation tools.

References

[1] Husmann P, O'Loughlin V 2018 *Anat. Sci. Educ*. **12** 6–19
[2] Newton P 2015 *Front. Psychol.* **6** 491–5
[3] Stowe K, Clinebell S 2016 *J. Int. Bus.* **26** 258–72
[4] Kraemer D, Hamilton R, Messing S, DeSantis J, Thompson-Schill S 2014 *Front. Hum. Neurosci*. **8** 1–9
[5] Koć-Januchta M, Höffler T, Thoma G, Prechtl H, Leutner D 2017 *Comput. Hum. Behav*. C **68** 170–9
[6] Maidment C, Jones C, Webb T, Hathway E, Gilbertson J 2014 *Energy Policy* C **65** 583–93
[7] Kaplan H, Brady P, Dritz M, Hooper D, Linam W, Froehle C, Margolis, P 2010 *Milbank Q.* **88** 500–59
[8] Meli P, Rey Benayas J, Balvanera P, Martínez Ramos M 2014 *PLoS ONE* **9** 93507–10
[9] Kwon E, King K, Nyilasy G, Reid L 2019 *J. Advert. Res.* 59(1) 99–128
[10] Alkhasawneh I, Mrayyan M, Docherty C, Alashram S, Yousef H 2008 *Nurs. Educ. Today* **28** 572–9
[11] Smith S, Handy J 2014 *J. Exp. Psychol. Learn. Mem. Cogn.* **40** 1582–93
[12] Lee C-F, Chuang S-C, Chiu C-K, Lan K-H. 2017 *Curr. Psychol.* **36** 1–18
[13] Kunter M, Frenzel A, Nagy G, Baumert J, Pekrun R 2011 *Contemp. Educ. Psychol.* **36** 289–301
[14] Robie C, Schmit M, Ryan A, Zickar M. 2000 *Organ. Res. Methods* **3** 348–65
[15] Hartnett M, George A, Dron J 2011 *IRRODL* **12(6)** 20–38
[16] Bastalich W 2015 *Stud. High. Educ.* **42** 1145–57
[17] Forissier T, Bourdeau J, Mazabraud Y, Nkambou R 2013 *Context* **LNAI8175** 330–5
[18] Koens F, Mann K, Custers E, Cate Ten O 2005 *Med. Educ.* **39** 1243–9
[19] Parker J, Hollister D, Gonzalez A, Brézillon P, Parker S 2013 *Context***LNAI8175** 45–58
[20] Fleming N, Mills C. Not another inventory, rather a catalyst for reflection. 1992
[21] Leite W, Svinicki M, Shi Y 2010 *Educ. Psychol Meas.* **70** 323–39
[22] Wang W-C, Cheng Y-Y, Wilson M 2016 *Educ. Psychol. Meas.* **65** 5–27
[23] Fleming N. VARK: a Guide to Learning Styles. The VARK Questionnaire. 2001
[24] R Core Team. 2018 Available from: https://www.R-project.org/
[25] Chalmers R 2012 *J. Stat. Soft.* **48** 1–29
[26] Bonifay W 2014 *Handbook of Item Response Theory Modeling* Routlege
[27] Hair J Jr, Black W, Babin B, Anderson R 2009 *Multivariate Data Analysis* Pearson