ISSN 2630-516X (Online)
Volume 2 Issue 1 · March 2019

Journal of International Education and Practice

Bilingual Publishing Co.

Volume 2 Issue 1 • March 2019 • ISSN 2630-516X (Online)

Journal of International Education and Practice

BPC 双语出版社
BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Volume 2 | Issue 1 | March 2019 | Page 1-59
Journal of International Education and Practice

Contents

Article

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep

ARTICLE

Using Hierarchical Generalized Linear Modeling to Examine Contextual Differential Item Functioning: A Validity Study of Teachers' Sense of Efficacy Scale

Jing Zhao[1]* Xiaojing Zou[2] Wenpeng Shang[3]

1. Institutional Research, Old Dominion University, Norfolk, United States
2. Chinese and Culture College, Beijing Normal University, Beijing, China
3. College of Foreign Studies, Jinan University, Guangzhou, China

ARTICLE INFO

Article history:
Received: 10 December 2018
Accepted: 31 January 2019
Published: 22 February 2019

Keywords:
HGLM
Teachers' Sense of Efficacy Scale
DIF

ABSTRACT

The purpose of the study was to further investigate the validity of the instrument used for collecting preservice teachers' perceptions of self-efficacy adapting the three-level hierarchical generalized linear modeling (HGLM) model. To serve the purpose, the study used data collected by the research team which elicited preservice teachers' self-efficacy beliefs using *Teachers' Sense of Efficacy Scale (TSES)*. HGLM were used to analyze the data. Results of the HGLM analyses (at level-two) showed that one item in the scale displayed gender DIF. Another item became DIF item when the context variable was added to the level-two model. However, the effect of the context on the DIF item was not big.

1. Introduction

Differential Item Functioning (DIF)

Differential Item Functioning (DIF) analysis has been and will continue to be a popular topic in the field of measurement. DIF refers to differences in the statistical properties of an item between groups of equal ability. The presence of DIF items on a measurement instrument threatens validity of the interpretation of scores (Angoff, 1993).[2] A similar view was shared by Maller (2001)[15] in claiming that DIF items may pose a threat to validity of test scores and may have serious consequences for groups as well as individuals. The probabilities of getting an item correct are determined not only by the trait that the test claims to measure but also by factors specific to groups, such as differential opportunities to learn or differences in socialization.

There are many different ways to look for DIF, which include but are not limited to Mantel-Haenszel, logistic regression, standardization, and item response theory (IRT).

**Corresponding Author:*
Jing Zhao,
Associate Director, Institutional Research, Old Dominion University, Norfolk, VA 23529*, United States*
Email: kellyzjing@gmail.com.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

In recent years, the nested nature of measurement data, such as items within students, students within schools, has also necessitated the need to utilize multi-level hierarchical linear modeling models to detect DIF. The multilevel model has a lot of advantages, as it could provide more accurate estimates of the standard errors of the parameters in the model. As a result, this kind of approach allows the researchers to investigate the impact of different predictors in the higher level units (e.g., schools, curriculum, and classrooms) and on the lower level units (e.g., students) (Kamata, 2001).[11] Because of these advantages of multilevel models, many studies have been done on detecting DIF using multilevel models (e.g., Kamata, 2001; Lin et al., 2016; Williams & Beretvas, 2006).[11][13][23]These studies used two-level models to investigate DIF items.

As an extension to the two-level models, Cheong (2006)[5]used a three-level model to investigate ethnic-racial DIF for 13 dichotomously-scored items which assessed civic-related language skills. Participants in the study included 2,076 U.S. ninth-grade students in 92 schools. In this model, multiple responses to items on rating civic-related language skills (level 1) were nested within students (level 2) and students were nested within schools (level 3). The analysis for the three-level model consisted of four steps: 1) set up the unconditional model, 2) estimate a conditional model with group membership as predictors, 3) investigate and assess which items exhibit DIF and the patterns, directions, and magnitude of the detected DIF, and 4) include the school-level variable in the DIF screening procedure and repeat Steps 2 and 3.

According to Cheong (2006),[5] it was important to include the contextual sources of DIF into an analysis and the use of a multilevel structure has several advantages. For example, the three-level model could provide a framework which integrates measurement properties and structural relationships; therefore, the framework could be used by researchers to evaluate the psychometric properties of the surveys. It also enables the researchers to model the performance of the items and students with student-, classroom-, and school-level variables. As most large-scale assessment programs have the data with nested structure, the approach is very useful in reality.

Because of its advantages, the three-level model was used to investigate whether items in TSES displayed DIF. Cheong's study applied the three-level model to analyze dichotomously-scored items and the present study applied the model to polytomously-scored items.

Teacher Self-Efficacy Beliefs

In his seminal work, Self-efficacy: The exercise of control, Bandura (1997)[3] first termed self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Teachers' efficacy beliefs were defined by Tschannen-Moran et al. (1998)[22] as "teacher's belief in his or her own capability to organize and execute courses of action required to successfully accomplish a specific teaching task in a particular context" (p. 233). Tschannen-Moran and Wolfolk-Hoy (2001)[21] reported that teacher self-efficacy beliefs were related to student outcomes, such as achievement motivation, and the students' own sense of efficacy.
Many studies have focused on the gender differences on preservice teachers' self-efficacy beliefs. Liu's study (2008)[14] showed that female preservice teachers at the early childhood, middle childhood, and secondary licensure levels reported higher self-efficacy beliefs than their male counterparts. However, it was found in Gülten's study (2013)[7] that preservice primary mathematics teachers' math literacy self-efficacy beliefs showed no difference in terms of gender.

The majority of studies on preservice teachers' self-efficacy beliefs have simply compared mean scale differences between female and male preservice teachers, without knowing whether items in a teacher self-efficacy measure function equally across gender. Gender is one of the most important grouping variable for DIF detection for measurement in the domains of self-concept (Yin & Fan, 2003),[24] emotion regulation (Anderson et al., 2016),[1] and learning disabilities (Murray et al., 2015).[17] Only when a measure is invariant across groups, the inference of group differences based on the mean difference is valid (Drasgow, 1987).[6] Therefore, it is necessary to investigate whether items in the preservice teachers' self-efficacy survey display differential item functioning (DIF) across gender.

The purpose of the study was to further investigate the validity of instruments used for collecting preservice teachers' perceptions of self-efficacy adapting the three-level HGLM model described in Cheong's study (2006).[5] The focus of the present study was to investigate whether the polytomously-scored items on the preservice teachers' self-efficacy survey function the same across gender and within different school contexts. Hence, the present study also explored contextual DIF. The research questions were:

What are the psychometric properties (e.g., reliability, construct validity) of the TSES?

Do items on the TSES display DIF by gender?

What is the effect of the third level variable, school context (public universities versus private universities) on gender DIF?

DOI: https://doi.org/10.30564/jiep.v2i1.395

2. Methodology

2.1 Instruments

TSES was constructed by Tschannen-Moran and Woolfolk Hoy (2001)[21] and it has been widely used in measuring teachers' efficacy beliefs. The short form of TSES (see Appendix A) was adopted in the study. The short form of the TSES consists of 12 items which are divided into three factors: efficacy for classroom management (CM), efficacy for student engagement (SE), and efficacy for instructional strategies (IS). The 12 items are rated on a 9-point Likert-type scale ranging from 1 = "nothing" to 9 = "a great deal".

Data in the study were collected by The Ohio State University Teacher Quality Program (TQP) research team in 2007 and 2008. The TQP program was a research initiative that included 50 colleges and universities that provided teacher preparation programs in the State of Ohio. The short form of TSES was used by TQP to assess preservice teachers' self-efficacy beliefs. Preservice teachers included in the study were in the final year of their teacher preparation program and about to graduate.

2.2 Participants

Participants in the study were 1,485 preservice teachers in 45 colleges/universities in the State of Ohio who were pursuing licensure to teach high school in the years of 2006-2007 and 2007-2008. The present study focused on the two cohorts and it was a cross-sectional study. The present study focused on secondary preservice teachers as the number of the female and male was more balanced than elementary preservice teachers in the dataset with eight hundred seven secondary preservice teachers being female (54%) and six hundred seventy eight being male (46%). In addition, 887 were attending the program in public universities (60%) and 598 were in private universities (40%).

2.3 Data analysis

Confirmatory Factor Analysis (CFA)

Cronbach's alpha and confirmatory factor analysis were conducted to answer the first research question, what are the psychometric properties of the TSES. Structural equation modeling takes many factors into account such as: interactions of variables, measurement error, and multiple latent independent variables which could be measured by multiple indicators; therefore, it is more powerful than other statistical approaches such as linear multiple regression (Kline, 1998).[12] LISREL 8.7 (Jöreskog & Sörbom, 2004)[10] was used to conduct the CFA analysis.

Hierarchical Generalized Linear Modeling (HGLM)

A three-level HGLM model (at level two) was applied to answer research question 2, whether the items in the instrument display DIF across gender. To answer research question 3, a third level variable, institution (public vs private universities) was added to the two-level model to see whether the results change. The three-level model was estimated adopting the procedures described in Cheong (2006)[5] which consisted of item responses as Level-1 units, secondary preservice teachers as Level-2 units, and institutions (public vs private universities) as Level-3 units.

Step 1: Estimate the Unconditional Model.

At level 1, for response i for teacher j in school k to item p, the model is:

$$\eta_{mijk} = \pi_{0jk} + \sum_{p=2}^{P} \pi_{pjk} X_{pijk} + \sum_{m=2}^{M-1} D_{mij} \delta_m \quad (1)$$

where η_{mijk} is the log-odds of the probability for the i-th response in category m or lower for teacher j in school k. π_{0jk} are adjusted log-odds of a response on a typical item for teacher j in school k, X_{pijk} are predictor variable representing response i for teacher j in school k to item p, and π_{pjk} are the coefficients for predictors X_{pijk} .The first item in each subscale of the two instruments was used as the reference item. δ_m is a threshold that separates categories m-1 and m. Dmij is an indicator for category m.

At level 2,

$$\pi_{0jk} = \beta_{00k} + \mu_{0jk},\ \mu_{0jk} \sim \mathrm{N}(0, \tau) \quad (2)$$

$$\pi_{0jk} = \beta_{p0k} \text{ for } p = 2, \ldots \mathrm{P}$$

Where β_{00k} is the intercept for school k on the performance of self-efficacy beliefs; μ_{0jk}

is assumed to be normally distributed with mean zero and teacher-level variance τ; π_{0jk}

are assumed to be invariant over teachers.

At level 3,

$$\beta_{00k} = \gamma_{000} + v_{00k},\ v_{00k} \sim N(0, \omega_0) \quad (3)$$

$$\beta_{p0k} = \gamma_{p00} \text{ for } p = 2 \ldots \mathrm{P}$$

Where γ_{000} is the mean level of performance on self-efficacy beliefs; the

random effect, β_{p0k} is assumed to be invariant across universities; and v_{00k} is assumed to be normally distributed with a mean of zero and university-level variance ω.

Step 2: Estimate a Conditional Model and Assess Individual Items.

The conditional model at level 1 was the same as the level 1 unconditional model in equation 2. At level 2, the grouping variable is gender (female=0 and male=1). The level-2 model is:

 DOI: https://doi.org/10.30564/jiep.v2i1.395

$$\pi_{0jk} = \beta_{00k} + \beta_{0k}Group_{jk} + \mu_{0jk},\ \mu_{0jk} \sim N(0,\ \tau) \quad (4)$$
$$\pi_{pjk} = \beta_{p0k} + \beta_{pk}Group_{jk} \text{ for } p = 2 \ldots P$$

Where Groupjk is a predictor of membership of group (female=0 and male=1) for teacher j in school k, β_{00k} is the intercept for school k on the performance of self-efficacy beliefs, β_{0k} is the group difference in self-efficacy beliefs between the reference group and group g for school k, β_{p0k} is the item effect for school k, and β_{pk} is the difference in the item effect between the reference group and group k.

In the level-2 model, if an item p does not display DIF, the difference between two groups in their expected log-odds of correct responses to the item should depend solely on the differences in their levels of self-efficacy beliefs and be equal to zero. If the estimate for an item p is not equal to zero, then the item can be judged to display DIF.

At level 3,

$$\beta_{00k} = \gamma_{000} + v_{00k}, v_{00k} \sim N(0, \omega_0) \quad (5)$$
$$\beta_{0k} = \gamma_{00} + v_{0k}, v_{0k} \sim N(0, \omega_g)$$
$$\beta_{p0k} = \gamma_{p00}$$
$$\beta_{pk} = \gamma_{p0}$$

Where γ_{000} the mean level of performance on self-efficacy beliefs and γ_{00} is the difference in self-efficacy beliefs between the reference group and group g.

Step 3: Enter a School-Level Correlate.

In the level-3 model, institution (public universities vs private universities) was added to the level-two model to see whether the school context has any effect on gender DIF. The level-three model is:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}Z_j + v_{00k}, v_{00k} \sim N(0, \omega_0) \quad (6)$$
$$\beta_{0k} = \gamma_{00} + v_{0k}, v_{0k} \sim N(0, \omega_g)$$
$$\beta_{p0k} = \gamma_{p00} + \gamma_{p01}Z_j$$
$$\beta_{pk} = \gamma_{p0} \text{ for } p = 2, \ldots P$$

where β_{00k} is the intercept for school k, β_{0k} is the group difference in self-efficacy beliefs for school k, β_{p0k} is the item effect for school k, β_{pk} is the difference in the item effect between group g and the reference group. γ_{000} is the grand mean level of performance on self-efficacy beliefs and γ_{00} is the difference in self-efficacy beliefs between the reference group and group g. Zj is the school context (public = 0 and private = 1) of teacher j.

For the level-3 model, the logic of detecting DIF for the level-2 model also applies. But for the level-3 model, the group difference in the item difficulties will be adjusted for various levels of the context variables. The multilevel models were run using HLM 6.04 (Raudenbush et al., 2004)[18] with the full penalized quasi likelihood (PQL). Missing data were handled while running the multilevel analyses.

3. Results

3.1 Descriptive Statistics

Descriptive statistics for TSES items were calculated using SPSS 19.0 and are displayed in Table 1. As reflected in the table, the average responses to these items ranged from 6.73 to 7.97 on a 9-point scale. In general, the preservice teachers were pretty sure about their abilities in managing classroom, engaging students, and using instructional strategies. For the preservice teachers, question 5 got the most agreement (M=7.97, SD=1.17) and question 2 got the least agreement (M=6.73, SD=1.53). That is to say, the secondary preservice teachers were most confident about their ability to craft good questions for their students and they were less but still confident about their skills in motivating students who show low interests in school work.(See Table 1.)

3.2 Confirmatory Factor Analysis for TSES

To answer research question 1, Cronbach's alpha and a three-factor CFA model were run for the instrument. Usually normal theory weighted least square chi-square $\chi 2$, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and goodness of fit index (GFI) are used to assess whether a model is a good fit to the data. Since the $\chi 2$ test is sensitive to the sample size, Shumacker and Lomax (2004)[20] suggest using RMSEA, SRMR, and GFI as indices to test the model fit. Browne and Cudeck (1993)[4]- suggest a RMSEA value between .05 and .08 for a good CFA model fit. According to Hu and Bentler (1999),[9] an SRMR value less than .08 and a GFI value greater than .95 mean a good fit.

The path diagram for the three-factor model is presented in Figure 1. As can be seen in Figure 1, for the latent variable, one estimated factor loading for each of the three–factors (i.e., CM, SE, IS) was fixed to 1. All of the 12 observed variables had significant factor loadings (ranging from .65 to .83), ps < .05, on the three factors. Hence, between 42% and 69% of the variance in teacher self-efficacy beliefs can be attributed to the three subscales and the overall rating. Teacher self-efficacy beliefs were appropriately and adequately assessed by the 12 observed variables in the TSES instrument. For the three-factor model, $\chi 2 = 444.27$, df= 51, p < .001, RMSEA = .072, SRMR = .038, and GFI =.95. Although $\chi 2$ is significant, it is due to the large sample size (Schumacker & Lomax, 2004).[20] The acceptable RMSEA, SRMR, and GFI values showed that the three-factor model was a good fit to the data. The CFA results also showed that conducting DIF analysis within each subscale of TSES would ensure that

DOI: https://doi.org/10.30564/jiep.v2i1.395

Table 1. Descriptive Statistics for TSES

		n	M	SD	Min.	Max.
Self-efficacy belief items						
CM	Control behavior (1)	1475	7.55	1.28	2	9
	Get students to follow rules (6)	1469	7.62	1.21	2	9
	Calm students (7)	1475	7.35	1.28	2	9
	Establish system (8)	1475	7.67	1.25	2	9
	Motivate students (2)	1476	6.73	1.53	1	9
SE	Get students to believe they can do well (3)	1473	7.53	1.21	3	9
	Help students value learning (4)	1472	7.15	1.37	2	9
	Assist families in helping students (11)	1469	6.76	1.56	1	9
	Craft questions (5)	1475	7.97	1.17	3	9
IS	Use assessment strategies (9)	1471	7.82	1.28	1	9
	Alternative explanation (10)	1470	7.90	1.16	2	9
	Alternative strategies (12)	1472	7.48	1.29	1	9
Preservice teachers gender						
	Male	1485	46%	50%	0	1
Institution						
	Private	45	40%	49%	0	1

the assumption of unidimensionality is met.

The reliability index for the overall scale, as well as reliability indices for the three subscales, as measured by Cronbach's alpha, was calculated using SPSS 19.0. The results showed that the reliability index for the overall scale is .905. The reliability index for the first factor, CM (items 1, 6, 7, and 8) is .880. The reliability index for the second factor, SE (items 2, 3, 4, and 11) is .817. The reliability index for the third factor, IS (items 5, 9, 10, and 12) is .802.

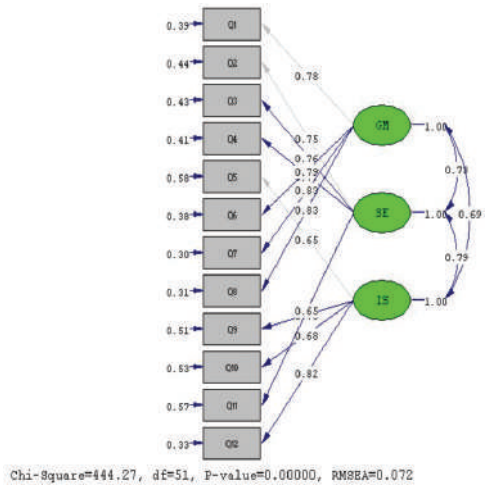


Figure 1. Path Diagram for TSES Three-Factor Model

A three-level HGLM model was used to answer research question 2, whether items in the instrument display gender DIF. To answer research question 3, what is the effect of the school context on gender DIF, institution was entered to the three-level models as a third-level correlate to see whether the results changed.

3.3 Multilevel analyses for TSES subscale 1

The results of the multilevel analyses for TSES subscale 1, CM, are presented in the following tables. For all the four items in this subscale, no teachers chose category 1 as responses. For the analysis, the maximum number of items was 5864, the maximum number of preservice teachers was 1466, and the maximum number of universities was 45.

Step 1: Estimate the Unconditional Model

The results of the unconditional model for subscale 1 are presented in Table 2. The first Thurstone threshold of the reference item, item 1 (How much can you do to control disruptive behavior in the classroom?) was -10.27. In addition, there was significant variance at both the teacher and university levels (6.02 and 0.25 respectively, $p<.001$). Therefore, predictors could be added to the model to account for the variance. Furthermore, distinguishing among teachers across university in self-efficacy beliefs can be

done with a pretty good reliability (0.85).

Table 2. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Unconditional Model for TSES Subscale 1

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.27	0.65	-15.79**
Item 6	γ_{100}	-0.18	0.09	-2.14*
Item 7	γ_{200}	0.48	0.11	4.49**
Item 8	γ_{300}	-0.33	0.06	-5.58**
Threshold 3	δ_3	2.42	0.56	4.36**
Threshold 4	δ_4	3.56	0.60	5.94**
Threshold 5	δ_5	5.65	0.63	9.00**
Threshold 6	δ_6	7.23	0.65	11.17**
Threshold 7	δ_7	9.93	0.65	15.36**
Threshold 8	δ_8	12.04	0.65	18.62**
	Random effects	Variance components		
Teacher-level variance	u_0	6.02**		
University-level variance	v_{00}	0.25**		

Note. df=44 for Intercept. df=5854 for items and thresholds. se=robust standard error. *p<.05. **p<.001.

Step 2: Estimate the Conditional Model and Investigate DIF Items

The results of the conditional model with preservice teachers' gender as the grouping variable are shown in Table 3. The coefficients for all the items in the subscale were not significantly different from zero with p>.05, therefore, all of the teacher self-efficacy belief items did not display significant gender DIF. Furthermore, the teacher-level variance for the model is significant (6.01, p<.001), which means that there is still a lot of variance in preservice teachers' self-belief efficacy beliefs, therefore, other predictors other than gender could be added to the model to account for the variance. The university-level variance is also significant (0.48 with p<.001), indicating that there is still a lot of variance to be explained and university level predictors such as institution could be added to the model to account for the variance.

Table 3. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting DIF for TSES subscale 1

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.26	0.65	-15.82**
TeachGender	γ_{010}	-0.07	0.15	-0.44
Item 6	γ_{100}	-0.27	0.12	-2.18*
Item 6 x TeachGender	γ_{110}	0.18	0.16	1.15
Item 7	γ_{200}	0.57	0.12	4.91**
Item 7 x TeachGender	γ_{210}	-0.19	0.12	-1.55
Item 8	γ_{300}	-0.43	0.09	-5.03**
Item 8 x TeachGender	γ_{310}	0.23	0.14	1.60
Threshold 3	δ_3	2.43	0.55	4.42**
Threshold 4	δ_4	3.57	0.59	6.03**
Threshold 5	δ_5	5.67	0.62	9.14**
Threshold 6	δ_6	7.25	0.64	11.35**
Threshold 7	δ_7	9.96	0.64	15.62**
Threshold 8	δ_8	12.07	0.64	18.95**
	Random effects	Variance Component		
Teacher-level variance	u_0	6.01**		
University-level variance	v_{00}	0.48**		
University-level gender variance	v_{01}	0.13		

Note. TeacherGender=teacher gender (0 = female and 1=male). df = 44 for Intercept and TeachGender. df = 5850 for items and the interaction of items and TeachGender, and thresholds.*p < .05. **p<.001.

Step 3: Enter a University-Level Correlate

The results for the conditional model with institution entered into the model are presented in Table 4. The coefficients for the interactions of institution and item 6 (How much can you do to get children to follow classroom rules? p<.05), of institution and item 7 (How much can you do to calm a student who is disruptive or noisy? p<.05), and institution and item 8 (How well can you establish a classroom management system with each group of students? p<.05) were significantly different from 0, therefore, statistically significant interactions were observed to exist between institution and these items. The positive coefficients for the three items indicated that preservice teachers at public universities had more

DOI: https://doi.org/10.30564/jiep.v2i1.395

confidence in getting children to follow classroom rules, in calming a student who is disruptive or noisy, and in establishing a classroom management system with each group of students. In addition, item 8 was found to become an item with gender DIF as the coefficient for this item was significantly different from zero ($p<.05$), which indicated that the context had some effects on gender DIF. The large p value (.04) also showed that the magnitude of DIF for this item was not big and the effect of the context on the DIF item was not big. The positive coefficient for item 8 indicated that female preservice teachers felt more confident in establishing a classroom management system than male preservice teachers. Furthermore, no significant interactions were observed by teacher gender and institution interactions. There were statistically significant between-university variations in self-efficacy beliefs, indicating that more predictors other than institution could be added to the model to account for the variance.

Table 4. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting Contextual DIF for TSES subscale 1

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.20	0.63	-16.12**
Insti	γ_{001}	-0.19	0.33	-0.59
TeachGender	γ_{010}	0.10	0.20	0.51
Insti x TeachGender	γ_{011}	-0.45	0.32	-1.41
Item 6	γ_{100}	-0.50	0.15	-3.32*
Item 6 x Insti	γ_{101}	0.54	0.21	2.60*
Item 6 xTeachGender	γ_{110}	0.27	0.21	1.24
Item 6 x InstixTeachGender	γ_{111}	-0.14	0.31	-0.46
Item 7	γ_{200}	0.36	0.15	2.44*
Item 7 x Insti	γ_{201}	0.49	0.23	2.14*
Item 7 xTeachGender	γ_{210}	-0.10	0.13	-0.82
Item 7 x Insti xTeachGender	γ_{211}	-0.17	0.29	-0.57
Item 8	γ_{300}	-0.58	0.08	-7.33**
Item 8 x Insti	γ_{301}	0.35	0.17	2.08*
Item 8 xTeachGender	γ_{310}	0.34	0.17	2.01*
Item 8 x Insti xTeachGender	γ_{311}	-0.25	0.29	-0.86
Threshold 3	δ_3	2 .44	0.54	4.49**
Threshold 4	δ_4	3.58	0.59	6.12**
Threshold 5	δ_5	5.68	0.61	9.27**
Threshold 6	δ_6	7.27	0.63	11.52**
Threshold 7	δ_7	9.98	0.63	15.86**
Threshold 8	δ_8	12.10	0.63	19.25**
	Random effects	Variance component		
Teacher-level variance	u_0	6.01**		
University-level variance	v_{00}	0.48**		
University-level gender variance	v_{01}	0.13		

Note. Insti=Institution (0=public university and 1=private university). df=43 for Intercept, TeachGender, and interaction of TeachGender and Institution. df=5842 for items and the interaction of items and TeachGender and Institution, and thresholds.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

3.4 Multilevel analyses for TSES subscale 2

For the analyses of TSES subscale 2, SE, the maximum number of items was 5844, the maximum number of preservice teachers was 1461, and the maximum number of universities was 45.

Step 1: Estimate the Unconditional Model

The results of the unconditional model for TES subscale 2, SE, are presented in Table 5. The first Thurstone threshold of the reference item, item 2 (How much can you do to motivate students who show low interest in school work?) was -8.53. In addition, there was significant variance at both the teacher (3.69, p<.001) and university level (.24, p<.001). Furthermore, distinguishing among teachers across university in self-efficacy beliefs can be done with high reliability (0.80).

Table 5. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Unconditional Model for TSES subscale 2

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-8.53	0.57	-14.84**
Item 3	γ_{100}	-1.45	0.07	-22.29**
Item 4	γ_{200}	-0.76	0.06	-12.64**
Item 11	γ_{300}	-0.10	0.06	-1.66
Threshold 2	δ_2	1.68	0.61	2.76*
Threshold 3	δ_3	3.54	0.55	6.46**
Threshold 4	δ_4	4.47	0.57	7.90**
Threshold 5	δ_5	6.29	0.59	10.64**
Threshold 6	δ_6	7.70	0.59	13.14**
Threshold 7	δ_7	9.67	0.59	16.27**
Threshold 8	δ_8	11.30	0.60	18.84**
	Random effects	Variance Component		
Teacher-level variance	u_0	3.69**		
University-level variance	v_{00}	0.24**		

Note. df=44 for Intercept. df=5833 for items and thresholds. se=robust standard error. *p<.05. **p<.001.

Step 2: Estimate the Conditional Model and Investigate DIF Items

The results for the conditional model with teachers' gender as the grouping variable are shown in Table 6. The coefficients for all the items in the subscale were not significantly different from zero with p>.05, therefore, all of the teacher self-efficacy belief items did not display significant gender DIF. Furthermore, the teacher-level variance for the model is significant (3.67, p<.001). The university-level variance is also significant (0.35, p<.001).

Table 6. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting DIF for TSES subscale 2

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-8.65	0.57	-15.13**
TeachGender	γ_{010}	0.25	0.14	1.76
Item 3	γ_{100}	-1.50	0.08	-18.16**
Item 3 xTeachGender	γ_{110}	0.10	0.12	0.83
Item 4	γ_{200}	-0.74	0.07	-11.33**
Item 4 xTeachGender	γ_{210}	-0.03	0.10	-0.28
Item 11	γ_{300}	-0.09	0.10	-0.92
Item 11 x TeachGender	γ_{310}	-0.01	0.16	-0.03
Threshold 2	δ_2	1.68	0.61	2.76*
Threshold 3	δ_3	3.54	0.55	6.47**
Threshold 4	δ_4	4.47	0.57	7.91**
Threshold 5	δ_5	6.29	0.59	10.65**
Threshold 6	δ_6	7.70	0.59	13.15**
Threshold 7	δ_7	9.68	0.59	16.29**
Threshold 8	δ_8	11.30	0.60	18.88**
	Random effects	Variance component		
Teacher-level variance	u_0	3.67**		
University-level variance	v_{00}	0.35**		
University-level gender variance	v_{01}	0.06		

Note. TeachGender=teacher gender (0=female and 1=male). df=44 for Intercept and TeachGender. df=5829 for items, the interaction of items and TeachGender, and thresholds. *p<0.05. **p<0.001.

Step 3: Enter a University-Level Correlate

The results of the conditional model with university type entered into the model are presented in Table 7. Significant interactions were observed between university type and item 11 (How much can you assist families in helping their children do well in school? p<.05). Negative coefficient for this item showed that preservice teachers at private universities were more confident about assist-

DOI: https://doi.org/10.30564/jiep.v2i1.395

ing families in helping their children do well in school. In addition, no significant interactions were observed by teacher gender and university type interactions with $p>.05$. Furthermore, there were statistically significant between-university variations in self-efficacy beliefs (3.68, $p<.001$), indicating that more predictors other than university type could be added to the model to account for the variance.

Table 7. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting Contextual DIF for TSES subscale 2

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-8.56	0.61	-13.97**
UniType	γ_{001}	-0.07	0.31	-0.24
TeachGender	γ_{010}	0.36	0.19	1.89
UniType x TeachGender	γ_{011}	-0.29	0.23	-0.91
Item 3	γ_{100}	-1.49	0.07	-21.09**
Item 3 x UniType	γ_{101}	-0.01	0.18	-0.06
Item 3 xTeachGender	γ_{110}	0.07	0.10	0.73
Item 3 x UniTypexTeachGender	γ_{111}	0.07	0.28	0.24
Item 4	γ_{200}	-0.71	0.09	-8.18**
Item 4 x UniType	γ_{201}	-0.08	0.14	-0.63
Item 4 xTeachGender	γ_{210}	-0.11	0.07	-1.68
Item 4 x UniTypexTeachGender	γ_{211}	0.21	0.24	0.87
Item 11 x UniType	γ_{301}	-0.39	0.20	-1.99*
Item 11 xTeachGender	γ_{310}	-0.09	0.18	-0.52
Item 11 x UniTypexTeachGender	γ_{311}	0.18	0.34	0.53
Threshold 2	δ_2	1.68	0.61	2.76*
Threshold 3	δ_3	3.54	0.55	6.47**
Threshold 4	δ_4	4.47	0.57	7.91**
Threshold 5	δ_5	6.30	0.59	10.66**
Threshold 6	δ_6	7.71	0.59	13.17**
Threshold 7	δ_7	9.69	0.59	16.32**
Threshold 8	δ_8	11.31	0.60	18.91**
	Random effects	Variance component		
Teacher-level variance	u_0	3.68**		
University-level variance	v_{00}	0.33**		
University-level gender variance	v_{01}	0.07		

Note. UniType = university type (0 = public university and 1 = private university). df = 43 for Intercept, TeachGender, and interaction of TeachGender and UniType. df = 5821 for items and the interaction of items and TeachGender and UniType, and thresholds. *$p<.05$. **$p<.001$.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

3.5 Multilevel Analyses for TSES Subscale 3

For TSES subscale 3, IS, the maximum number of items was 5864, the maximum number of preservice teachers was 1466, and the maximum number of universities was 45.

Step 1: Estimate the Unconditional Model

The results for the unconditional model for TESE subscale 3, IS, are presented in Table 8. The first Thurstone threshold of the reference item (To what extent can you craft good questions for your students?) was -10.14. In addition, there was significant variance at both the teacher and university levels (3.21 and 0.11 respectively, $p<.001$). Furthermore, distinguishing among teachers across university in self-efficacy beliefs can be done with decent reliability (0.77).

Table 8. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Unconditional Model for TSES subscale 3

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.14	1.22	-8.30**
Item 9	γ_{100}	0.29	0.07	4.15**
Item 10	γ_{200}	0.20	0.07	2.66*
Item 12	γ_{300}	1.05	0.12	9.05**
Threshold 2	δ_2	0.70	0.76	0.93
Threshold 3	δ_3	2.75	1.15	2.40*
Threshold 4	δ_4	3.75	1.20	3.12*
Threshold 5	δ_5	5.76	1.17	4.94**
Threshold 6	δ_6	6.96	1.20	5.81**
Threshold 7	δ_7	8.92	1.19	7.51**
Threshold 8	δ_8	10.60	1.19	8.90**
	Random effects	Variance component		
Teacher-level variance	u_0	3.21**		
University-level variance	v_{00}	0.11**		

Note. df=44 for Intercept. df=5853 for items and thresholds. se=robust standard error. *$p<.05$. **$p<.001$.

Step 2: Estimate the Conditional Model and Investigate DIF Items

The results of the conditional model with preservice teachers' gender as the grouping variable are shown in Table 9. As shown in the table, item 9 displayed significant gender DIF with $p<.001$ and the coefficient for males and females was significantly different from zero. For item 9, the positive coefficient indicated that male preservice teachers found the item harder to endorse and felt less confidence about using different strategies in classroom given the same level of ability as female preservice teachers. Furthermore, the teacher-level variance for the model is significant (3.21, $p<.001$). The university-level variance in the model is also significant (0.13, $p<.05$).

Table 9. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting DIF for TSES subscale 3

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.24	1.23	-8.32**
TeachGender	γ_{010}	0.18	0.10	1.81
Item 9	γ_{100}	0.00	0.10	0.02
Item 9 xTeachGender	γ_{110}	0.62	0.12	5.13**
Item 10	γ_{200}	0.15	0.11	1.35
Item 10 xTeachGender	γ_{210}	0.10	0.16	0.63
Item 12	γ_{300}	1.00	0.16	6.31**
Item 12 x TeachGender	γ_{310}	0.10	0.13	0.75
Threshold 2	δ_2	0.70	0.76	0.93
Threshold 3	δ_3	2.75	1.15	2.40*
Threshold 4	δ_4	3.75	1.20	3.13*
Threshold 5	δ_5	5.77	1.16	4.96**
Threshold 6	δ_6	6.98	1.20	5.83**
Threshold 7	δ_7	8.94	1.18	7.55**
Threshold 8	δ_8	10.62	1.19	8.94**
	Random effects	Variance component		
Teacher-level variance	u_0	3.21**		
University-level variance	v_{00}	0.13*		
University-level gender variance	v_{01}	0.01		

Note. TeachGender = teacher gender (0 = female and 1 = male). df=44 for Intercept and TeachGender. df = 5849 for items , the interaction of items and TeachGender, and thresholds. *$p<.05$. **$p<.001$.

Step 3: Enter a University-Level Correlate

The results of the conditional model with university type entered into the model are presented in Table 10. Item 9 was still found to display gender DIF ($p<.001$) after institution was entered into the model. The context did

DOI: https://doi.org/10.30564/jiep.v2i1.395

not have any effect on the DIF item. No statistically significant interactions were observed between institution and preservice teachers' self-efficacy belief items with the coefficients being not significantly different from zero. In addition, no significant interactions were observed by teacher gender and institution interactions. Furthermore, there were statistically significant between-university variations in self-efficacy beliefs.

Table 10. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting Contextual DIF for TSES subscale 3

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.21	1.29	-7.93**
UniType	γ_{001}	-0.05	0.21	-0.23
TeachGender	γ_{010}	0.20	0.11	1.74
UniType x TeachGender	γ_{011}	-0.04	0.22	-0.20
Item 9	γ_{100}	0.01	0.13	0.11
Item 9 x UniType	γ_{101}	-0.03	0.20	-0.14
Item 9 xTeachGender	γ_{110}	0.57	0.14	4.04**
Item 9 x UniTypexTeachGender	γ_{111}	0.15	0.26	0.56
Item 10	γ_{200}	0.15	0.16	0.91
Item 10 x UniType	γ_{201}	0.01	0.22	0.04
Item 10 xTeachGender	γ_{210}	0.07	0.23	0.32
Item 10 x UniTypexTeachGender	γ_{211}	0.08	0.30	0.26
Item 12	γ_{300}	1.06	0.25	4.27**
Item 12 x UniType	γ_{301}	-0.13	0.30	-0.42
Item 12 xTeachGender	γ_{310}	0.14	0.19	0.72
Item 12 x UniTypexTeachGender	γ_{311}	-0.11	0.26	-0.42
Threshold 2	δ_2	0.70	0.75	0.93
Threshold 3	δ_3	2.75	1.14	2.41*
Threshold 4	δ_4	3.75	1.20	3.13*
Threshold 5	δ_5	5.77	1.16	4.97**
Threshold 6	δ_6	6.98	1.19	5.85**
Threshold 7	δ_7	8.94	1.18	7.57**
Threshold 8	δ_8	10.62	1.18	8.97**
	Random effects	Variance component		
Teacher-level variance	u_0	3.22**		
University-level variance	v_{00}	0.12*		
University-level gender variance	v_{01}	0.01		

Note. UniType = university type (0 = public university and 1 = private university). df = 43 for Intercept, TeachGender, and interaction of TeachGender and UniType. df = 5841 for items and the interaction of items and TeachGender and UniType. *p<.05. **p<.001.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

4. Discussion

Several findings emerged from the present study. As indicated by the descriptive statistics, the secondary preservice teachers in the present study were confident about their abilities in managing classroom, engaging students, and using instructional strategies. The preservice teachers were very confident about their ability inside the classroom, such as crafting good questions for their students and redirecting a student who was disruptive and noisy in the classroom. In addition, they were confident about their ability to motivate students who showed low interests in school work although a student's motivation and performance also depended on his or her home environment. There is some support for the claim that home environment was important for students' motivation. For example, Muola's study (2010)[16] showed that student's motivation to do well in their academic work is to a certain extent dependent on the nature of their home environment.

Results of the CFA analyses provided evidence that TSES had good construct validity. Results showed that the three-factor model for TSES worked well for the present sample in the study. In addition, reliabilities for the overall scales and each subscale are high, which shows that the instrument is a reliable measure of preservice teachers' self-efficacy beliefs.

In terms of DIF items, the results of multilevel HGLM analyses at level-two showed that for TSES subscales 1 (CM) and 2 (SE), no DIF items were detected. For subscale 3 (instructional strategy), item 9 (How much can you use a variety of assessment strategies?) was found to display significant gender DIF. Female secondary preservice teachers reported to be more confident about implementing alternative strategies in the classroom compared to male preservice teachers with the same ability. The reason might be that female preservice teachers practiced using more alternative strategies than male preservice teachers. Further follow-up such as content analysis should be conducted to see whether this item is a biased item. If it is a biased item, then it should be modified or deleted from the instrument.

Item 6 (How much can you do to get children to follow classroom rules?), item 7 (How much can you do to calm a student who is disruptive or noisy?) and item 8 (How well can you establish a classroom management system with each group of students?) in TSES 1were found to have significant interactions with institution. The positive coefficients for the three items showed that secondary preservice teachers who studied to get their licensures at public universities tended to be more confident about "getting children to follow classroom rules", "calming a disruptive or noisy student", and "establishing a classroom management system with each group of students" than their counterparts at private universities. The possible reason is that the teacher education programs at public universities in the State of Ohio were more reputable, therefore, preservice teachers at these universities were more confident about their teaching abilities.

The results of the multilevel analysis also showed that context had effect on DIF items, however, the effect was not big. Some studies have discussed the difference in teacher preparation programs at public versus private universities. For example, Henry et al. (2011)[8] investigated whether teachers got their licensures at public or private universities influenced the achievement of their students. Rosas and West (2011)[19] examined the perceptions of preservice teachers at both public and private universities in the State of Ohio regarding their readiness to teach. However, no research was found on why getting licensures at public or private universities (context) could have influence on preservice teachers' self-efficacy beliefs.

The present study, in the methodological sense, tests the applicability of the three-level model of Cheong (2006) [5] to polytomous data which are more often used in the field of education and psychology. In addition, the present study takes the contextual sources of DIF into account to see whether the context exerts any impact on gender DIF as the context might be one of the sources for DIF items (Cheong, 2006).[5] This multilevel approach has the advantages as evidenced in Cheong's study (2006)[5] as described earlier. The present study also contributes to the self-efficacy literature in that it further explores validity evidence for the widely used self-efficacy survey (TSES) from a new perspective, item bias, which has not been considered before in self-efficacy studies.

One limitation for the study is that due to the unique demographic characteristics of the participants, caution should be exercised when generalizing the results to the preservice teachers at large. The participants in the study were from the colleges and universities which provided licensure programs to preservice teachers in the State of Ohio. All of the participants in the present study were in bachelor's degree program. Around half of them were female and half were male. In addition, half of the preservice teachers were studying at public universities and half at private universities to pursue their licensures. The findings of the study can only be generalized to the population with similar characteristics.

5. Conclusion

To sum up, the present study aimed to investigate validity related issues of the TSES, which measured teachers' self-efficacy beliefs. 1485 preservice teachers who were pursuing their licensures in the colleges and universities

  DOI: https://doi.org/10.30564/jiep.v2i1.395

in the State of Ohio in the years of 2006-2007 and 2007-2008 participated in the present study.

The results of the descriptive statistics showed that the preservice teachers were pretty confident about their teaching ability in classrooms, but not so confident about how to motivate students. The results of the CFA analyses showed that the TSES has good construct validity and is a reliable measure for the present sample. The instrument could measure what it is supposed to measure.

In terms of DIF items, one item (item 9) was detected by multilevel model to be DIF items in TSES subscale 3, IS. In addition, item 8 in TSES subscale 1became DIF items when the context variable, institution, was added to the level-two model. Further follow up analysis can be done to investigate whether these items are biased item. Modification could be made to the item or the item could be deleted from the instrument. In general, the four-step procedure (modified to a three-step procedure in the present study) in Cheong's study (2006)[5] worked well for the present sample with high estimate reliability. However, due to some limitations of the present study, caution should be used in generalizing the results.

5.1 Implications for Future Research

The present study is a study using several quantitative techniques, CFA and HGLM analyses to further validate the TSES. The present study has its own significance as it contributes to the sparse literature investigating the validity issues related to measuring preservice teachers' self-efficacy beliefs. The present study also tests the applicability of the three-level model of Cheong (2006)[5] to polymotous data which are more often used in the field of education. However, it also has its limitations as discussed earlier. Therefore, further studies should be done in light of the limitations.

Due to the unique characteristics of the participants in the present study, additional studies using a randomly selected sample from diversified regions are needed to generalize the results. In addition, effect size was not reported in the present study. As effect size is a measure of practical significance and indicates the magnitude of the difference, therefore, in future studies, effect size should be included as part of the results of multi-level model. Also because this is only an empirical study which produced preliminary results with the present sample size, additional simulation studies which simulate different sample size and compare different procedures of DIF estimates should be conducted to have a more holistic result.

In the multilevel analyses, after teacher gender and institution were added to the model, there was still significant variance at the teacher and university level that needed to be explained, which suggested that more predictors could be added to the model. Therefore, further studies could add teacher-level predictors such as teachers' race and university level predictors such as mean SES to the analyses to see whether these variables could help account for the variance in the model. In addition, the results of the studies also showed that the school context had some effect on gender DIF. After the third-level variable (institutions) was added to the second level model, one item became DIF items from DIF-free items. Further studies could be done to show how the context (public universities and private universities) where preservice teachers attend teaching programs influences their self-efficacy beliefs.

Acknowledgement: We thank Dr. Dorinda J. Gallant, The Ohio State University, for her specific instructions and for the time and energy she put in the manuscript.

References

[1] Anderson, L. M., Reilly, E. E , Gorrell, S., Schaumberg, K., & Anderson, D. A. (2016). Gender-based differential item function for the difficulties in emotion regulation scale. Personality and Individual Differences, 92, 87-91.

[2] Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P.W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 3-24). Hillsdale, NJ: Lawrence Erlbaum.

[3] Bandura, A. (1997). Self-efficacy: The exercise of control. New York, NY: W. H. Freeman and Company.

[4] Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen and J. Long (Eds), Testing structural equation modeling (pp.136–62). Newbury Park, CA: Sage.

[5] Cheong, Y. F. (2006). Analysis of school context effects on differential item functioning using hierarchical generalized linear models. International Journal of Testing, 6, 57-79.

[6] Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. Journal of Applied Psychology, 72, 19-29.

[7] Gülten, D. Ç. (2013). An investigation of pre-service primary mathematics teachers' math literacy self-efficacy beliefs in terms of certain variables. International Online Journal of Educational Sciences, 5 (2), 293-408.

[8] Henry, G.T., Thompson, C.L., Bastian, K.C., Fortner, C.K. Kershaw, D.C., Purtell, K.M., & Zulli, R.A. (2011). Does teacher preparation affect student achievement? Manuscripts submitted for Education Finance and Policy.

[9] Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indices in covariance structure analysis:Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1–55.

[10] Jöreskog, K.G., & Sörbom, D. (2004). LISREL 8.7. Scientific Software International, Inc., 1981-2004, IL, U.S.A.

[11] Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. Journal of Educational Measurement, 38, 79-93.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

[12] Kline, R. B. (1998). Principles and practices of structural equation modeling. New York, NY: The Guilford Press.

[13] Lin, P.Y., Childs, R.A., & Lin, YC (2016). Untangling complex effects of disabilities and accommodations within a multilevel IRT framework. Quality & Quantity, 50, 2767-2788.

[14] Liu, J. (2008). A comparison of teacher candidates and first-year teachers by gender and licensure level, in terms of their perceptions of preparation program quality, efficacy beliefs, and concerns about teaching. Unpublished doctoral dissertation, The Ohio State University, USA.

[15] Maller, S. J. (2001). Differential item functioning in the WISC-III: Item parameters for boys and girls in the national standardization sample. Educational and Psychological Measurement, 61, 793-817.

[16] Muola, J.M. (2010). A study of the relationship between academic achievement motivation and home environment among standard eight pupils. Educational Research and Reviews, 5 (5), 213-217.

[17] Murray, A. L., Booth, T., & Mckenzie, K. (2015). An analysis of differential item functioning by gender in the Learning Disability Screening Questionnaire (LDSQ). Research in Developmental Disabilities, 39, 76-82.

[18] Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., du Toit, M. (2004). HLM 6:Hierarchical Linear and Nonlinear Modeling. Lincolnwood: Scientific Software International.

[19] Rosas, C. & West, M. (2011). Pre-service teachers' perception and beliefs of readiness to teach mathematics. Current Issues in Education, 14(1). Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/view/

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

[21] Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. Teaching and Teacher Education, 17, 783-805.

[22] Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. Review of Educational Research, 68(2), 202-248.

[23] Williams, N.J., & Beretvas, S. N. (2006). DIF identification using HGLM for polytomous items. Applied Psychological Measurement, 30, 22-42.

[24] Yin, P., & Fan, X. (2003). Assessing the factor structure invariance of self-concept measurement across ethnic and gender groups: Findings from a national sample. Educational and Psychological Measurement, 63(2), 296–318.

Appendix A: Teachers' Sense of Efficacy Scale (TSES)

Directions: This questionnaire is designed to help us gain a better understanding of the kinds of things that create difficulties for teachers in their school activities. On a 9 point scale where 1 is "nothing" and 9 is "a great deal," please mark the option corresponding to your opinion about each of the statements below.

	Nothing	2	Very Little	4	Some Influence	6	Quite a Bit	8	A Great Deal
1. How much can you do to control disruptive behavior in the classroom?	m	m	m	m	m	m	m	m	m
2. How much can you do to motivate students who show low interest in school work?	m	m	m	m	m	m	m	m	m
3. How much can you do to get students to believe they can do well in school work?	m	m	m	m	m	m	m	m	m
4. How much can you do to help your students value learning?	m	m	m	m	m	m	m	m	m
5. To what extent can you craft good questions for your students?	m	m	m	m	m	m	m	m	m
6. How much can you do to get children to follow classroom rules?	m	m	m	m	m	m	m	m	m
7. How much can you do to calm a student who is disruptive or noisy?	m	m	m	m	m	m	m	m	m
8. How well can you establish a classroom management system with each group of students?	m	m	m	m	m	m	m	m	m
9. How much can you use a variety of assessment strategies?	m	m	m	m	m	m	m	m	m
10. To what extent can you provide an alternative explanation or example when students are confused?	m	m	m	m	m	m	m	m	m
11. How much can you assist families in helping their children do well in school?	m	m	m	m	m	m	m	m	m
12. How well can you implement alternative strategies in your classroom?	m	m	m	m	m	m	m	m	m

DOI: https://doi.org/10.30564/jiep.v2i1.395

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep

ARTICLE

Integral Thinking and its application to Integral Education

Josep Gallifa*

FPCEE Blanquerna. Ramon Llull University, C. Cister 34, 08022 Barcelona, Spain

ARTICLE INFO

Article history
Received: 6 March 2019
Accepted: 13 March 2019
Published: 27 March 2019

Keywords:
Thinking skills
Integral thinking
Holonic theory
Part-whole relationships
Integral theory
Integral education

ABSTRACT

After the exploration of different kind of thinking skills oriented to discerning complex issues and phenomena, this article argues about the need to use a new modality of thinking, defined as integral thinking. Based on the holonic theory and on the proprieties of the holons and their part-whole relationships, the paper proposes and characterizes integral thinking, a kind of thinking that is holistic but also has span and profundity, going beyond conventional Aristotelian epistemic-inspired ways. Integral thinking joins in a single model the assets of premodern, modern and postmodern thinking systems; and helps in going beyond the nowadays syncretism of many perspectives. Integral thinking is the kind of thinking appropriated to the contemporary need to think integrally in science, culture, professions, and arts or about the evolution of personal consciousness. It's useful also to be applied in the diverse professional fields, especially when comprehensive approaches are needed. Integral thinking can be used in holistic education and pedagogies. The use of integral thinking in the educational *actuality* can help to characterize the integral education practice and agenda.

1. Introduction

In the information age in which we live is especially pertinent understanding thinking skills and their relevance in the systems of thought. Thinking skills are required in all the applied fields like education, psychotherapy, leadership, and organizations, among many other ones, as well as in the acquisition of personal consciousness. Today there is an increasing need for thinking well, an authentic art. We use the word 'art', from the Latin *ars*, because thinking can be understood also as a *tekhne*, the Aristotelian Greek word for art. Effectively Aristotle[1] proposed three kinds of *episteme*: *episteme theoretike, poietike, and praktike*. Each one has intellectual character[2] and represents, therefore, a kind of thinking. The first one allows thinkers to create theories, which conduct necessary outcomes, as is the case of theoretical predictive sciences. The second one allows the *poiesis* of different possibilities, which are separated from the thinker too. The *episteme praktike*, instead, is oriented to the *phronesis* and centered in the agent, in order to guide the *praxis*, the ways of acting. Aristotle introduced also the *nous*, "that grasps die self-identical essences which are objects of first philosophy"[3], involving "the complete identity of knower and known, which represents the ideal to which human noetic activity aspires"[4]. All those modalities together form, in his system, the *Sophia*. "On Aristotle's view, *Sophia* is a combination of *nous* and *episteme*"[5], which requires, therefore, the contribution of different modalities of thinking. The combination of these modalities can be

**Corresponding Author:*
Josep Gallifa,
Full Professor and Department Head, Faculty of Psychology, Education and Sport Sciences Blanquerna, Ramon Llull University, C. Cister 34, 08022 Barcelona, Spain;
E-mail: josepgr@blanquerna.url.edu.

 DOI: https://doi.org/10.30564/jiep.v2i1.603

described as wisdom[6]. A challenge for contemporary Psychology is how to define a system able to include all the Aristotelian thinking modalities: the more conventional *epistemes*, but especially the less treated *nous* and *Sophia*. This is the first purpose of this study.

It's a very relevant issue because Modernity, commonly considered, promoted rationality in the generalized *episteme theoretike* style of thinking. Modern schools and universities trained students in this kind of thinking, the one that De Bono[7] summarized in the sentence: "I'm right, you are wrong". Other concomitant characteristics of the modern worldview (like specialization, the predominance of analysis, fragmentation in disciplines, etc.), maybe helped to promote real progress, but couldn't encompass the total complexity of the human experience. To approach more complex views, disciplines like phenomenology, for instance, have to include in their methodological process the phenomenological reduction[8], in order to allow the possibility of thinking about subjective matters without being influenced too much by the established conventional theoretical ways.

Additionally, problems today are multifaceted, cross-disciplinary, human-centered, and rarely solved through simple or linear solutions[9] and for that reason, we need in our time many different kinds of thinking to deal with different kind of problems and phenomena. For instance, de Bono, aware of the western predominance of logic-formal thinking, presented six flexible modalities in his well-known model of the six thinking hats[10]: logic (black), emotional (red), creative (green), factual (white), positive (yellow), procedural (blue). Thus, another question may arise: Are there, in this perspective, all the thinking modalities represented? Our point of view is that, although this approach is useful in the kind of thinking needed in the creative arena of problem-solving, it doesn't exhaust all the possible options of thinking.

In order to complete the map of thinking modalities, and on the other hand, there is a new intellectual need in our postmodern societies: thinking about complexity[11]. That means to keep in mind at the same time diverse features: subjective-objective, biological-cultural, scientific-professional, and so on. New ways of thinking, going beyond unidirectional modalities, have to be proposed to create comprehensive systems of thought about complex situations. In that sense, Wilber[12] proposed the integral theory as a map for the different dimensions of human consciousness. We can inquire if there are corresponding modalities of thinking to the integral approach. Especially interesting would be to find out what are the modalities available in the advanced stages of consciousness. And because the integral theory has experimented a very generalized application in different fields[13], understanding the new thinking skills associated with it arises as a very relevant issue.

Wilber[14] additionally advocated for the need that a significant number of people arrive at the second tier of conscience. If this number would reach the 10% of the human population, Wilber explains that there would be a no return point and the humanity, as a collective, would make a big change in the collective consciousness and values[15]. If the styles of thinking taught at schools are based in the modern and in the Aristotelian thinking modalities is not strange that nowadays only maximum the 5% of humanity arrives at these advanced stages of consciousness[16]. And with only positive, wishful and creative thinking, the preferred postmodern modalities, there will be no significant advance. It's thus necessary to characterize a new modality of thinking to be used in advanced stages. Another purpose of this study is how to characterize this advanced modality of thinking independently of any religious or philosophical tradition, nevertheless with the potential to be recognizable for the different traditions, as a meeting point, where diverse people could meet and share advanced systems and models about the human consciousness and experience[17].

2. Descriptive Characterization of Integral Thinking

2.1 Advanced Modalities of Thinking

A list of advanced modalities of thinking, ordered by correspondences in the spectrum of consciousness[18], includes:

Logic-formal thinking. Piaget[19] studied and described operational-formal thinking, and documented its development. This kind of thinking makes use of mental operations, which are interiorized actions, and is applicable to solve mathematical and scientific problems.

Mythic-religious-cultural thinking. Religious and philosophical premodern traditions developed systems of thought around the relationship of the individual with the cosmos. Worldviews (cosmo-visions) are implicit in all human cultures. Each culture has its own belief systems, which often include the mythological sphere[20]. The collective cultural representation allowed massive cooperation and was necessary for the development of larger societies, developed from the original survival clans. These representations use imaginative human capabilities[21].

Rational, dialectic, rhetoric thinking. Aristotle, as has been explained, systematized different kinds of thinking. On one side the rational-dialectic thinking. Formal forums were developed, and western thought grew in different

DOI: https://doi.org/10.30564/jiep.v2i1.603

fields by exercising rationality: science, law, medicine, organizations, ethics, aesthetics, and so on[22]. The new psychological paradigm for the study of reasoning integrates psychology of reasoning with the study of judgment and decision-making, including degrees of uncertainty. Subjective psychological value or utility and social pragmatics are playing a central role[23]. On the other hand, Aristotle[24] systematized as well the *tekhne retorike*, the rhetoric thinking: An art of persuasion, will-mobilizing, employed to produce eloquent discourses. The purpose of this art is different from dialectics but utilizes also reasoning abilities.

Design-oriented, creative, positive thinking. Gardner[25] established creative lines of development parallel to the Piagetian logic-mathematical one (spatial, musical, linguistic, kinesthetic). In turn, De Bono[26] presented the creative and positive modalities of thinking skills, as an alternative to logic-argumentative kind of thinking. On the other hand, Bereiter & Scardamalia[27] systematized design thinking, which is proposal oriented, prototype developer, and socio-critical based. There is a dynamic relationship between thinking modalities and mental states. For example, creative thinking can be influenced by mindfulness meditation[28].

Symbolic-spiritual thinking. Jung[29] introduced the concept of symbol distinct from the concept of sign. While a sign is like an instrumental tool of the mind in order to communicate well-known realities, a symbol canalizes psychical energy from the Jungian collective unconscious. The brain, in the Jungian view, not only processes signs but also has symbolic capabilities. The symbolic thinking is necessary for evolution, from literal-mythic-cultural systems to a more personal, experiential kind of religiosity-mysticism. Gardner[30] documented inter and intrapersonal developmental lines, that use symbolic competences, especially the intrapersonal one.

Post-formal, vision-logic thinking. Sinnot and Commons & Richards[31] proposed postformal thinking, as subsequent stages after the Piagetian. Botella and Gallifa[32] described this postformal thinking as contextual, relativist, organicist and studied their contextual dependence. Wilber[33] named this kind of thinking as vision-logic because unites logical with symbolic capabilities. This thinking depends on experience and has his own development: Systemic, metasystemic, paradigmatic, meta-paradigmatic[34]. Intuition, in advanced modalities of thinking, has an impact at the moment of taking complex decisions[35].

Thinking of the "branches", spiritual advanced, metaphysical thinking. Ashlag[36], a Jewish Kabbalah rabbi, proposed the language of the "branches" to talk about metaphysical concepts. As it where a tree, some complex metaphysical concepts belong to the "roots" but, in order to talk about them, is necessary to develop a special language of the "branches" in order to share knowledge. This kind of thinking is necessary to understand some metaphysical concepts and is difficult to be separated from metaphysical content. It is the kind of thinking used in the relationship between a spiritual teacher and his/her disciple.

Integral thinking. Sharing traits with symbolic, postformal, and metaphysical modalities of thinking, there is the option of a kind of thinking aimed to amplify and concrete the possibility to think about complex matters, in the scientific, professional, cultural or personal arenas, without metaphysical reference. We'll describe and characterize it with the name of integral thinking.

2.2 The Need for Integral Thinking

The last years have seen "leading thinkers in many fields of scholarly endeavor (including complexity science, ecology, education, futures studies, integral studies, philosophy, psychology, spirituality studies, and systems theory) claim that the fragmented, mechanistic and materialistic ways of thinking of the last century are no longer sustainable. As Einstein put it a century ago, the significant problems we have cannot be solved at the same level of thinking with which we created them". [37]

The aforementioned need for having a kind of thinking aimed at thinking about complex issues is another argument to justify the demand for new ways of thinking. There is a need to develop "attempts to open futures thinking out in all possible dimensions to embrace the new thinking and knowledge patterns that are emerging across the breadth and depth of the global knowledge terrain".[38]

In that sense, Wilber[39] formulated the already mentioned integral theory of human consciousness and evolution. This theory joins in a single model the premodern, modern and postmodern legacies. The theory defines four quadrants covering subjective, objective, inter-subjective, and inter-objective fields of human experience. In order to interact intellectually with those fields, which are operating at the same time, and especially for the elevated levels of consciousness, we need to characterize a new modality of thinking: the integral thinking. Although the integral theory has been fully studied as a map for thinking[40], there has not been developed yet a system of thinking according to the need to think in the framework of this knowledge.

In order to think globally, there is the designated as 'holistic thinking'[41] (for example in holistic medicine), which promotes thinking about the whole human being instead of thinking about isolated parts. "The whole is greater than the sum of its parts", affirmed Aristotle[42].

 DOI: https://doi.org/10.30564/jiep.v2i1.603

But the relation between the 'parts' and the 'whole' will need more detailed exploration. Which are the parts? What is the relation between a single part and the whole? In which conditions one part can be at the same time part and whole? Is there any relationship between a part of a part and the corresponding part of the whole? All these questions are unresolved in generic 'holistic thinking'. Holistic thinking can be another way to name the firsts stages of vision-logic or post-formal thinking, inside the Aristotelian epistemic framework. To approach all these questions and to surpass the Aristotelian conventional-episteme-related views there is a need for integral ways of thinking.

2.3 What Does 'Integral' Mean?

'Integral', coming from the Latin '*integralis*', signifies "composed of parts that together constitute a whole"[43], and/or that all the parts are necessary for making the whole entire and complete[44]. There is another trait when a part is denominated as integral: It means that it's a necessary, essential or fundamental part[45]. A general trait of an integral approach can be found by joining the diverse meanings: Because the whole has to be complete, it requires an equilibrated, balanced, and irreplaceable presence of the constituent and essential parts.

As a verb "to integrate" could mean "to make whole or to make complete; to make part of a larger unit; to join with something else; to unite"[46], especially from similarities. The word "integrated" means composed of parts. "Integer", with reference to a person or a system, means entire (comprehensive), literally or figuratively, complete in itself.[47] In this last case could mean with a "character" or adhesion to a set of principles. When we use "integral" as an adjective the meaning is wider than "integrated" or "integer". The essential parts have to be balanced to create a united entire whole.

"An integral approach is inclusive and does not privilege particular parts over others; rather, people judiciously and with careful deliberation fuse relevant parts into new entities so as to address the complexity of the situation"[48]. The combination of the parts will be able "to get a new whole that provides the complexity required to address the unique situation"[49]. The resultant whole will be an emerging outcome. "By using the term integral, we foreground concepts of inclusivity, holism, pluralism, and reverence"[50].

It's the case of integral thinking as a modality of thinking about complete, comprehensive and balanced relations in part-whole issues. We need this kind of thinking because it is necessary to distinguish between an integrated approach (syncretism of different perspectives) from an integral one, as we will explain.

Integral thinking is useful to think about complexity. Complexity is composed of two independent dimensions:[51] differentiation and integration. Wilber[52] considered complexity as a continuing process of differentiating and integrating. Although these capabilities are necessary for complex thinking they are not sufficient to characterize the thinking about comprehensive complex realities. For that reason too, we are introducing integral thinking.

2.4 Definition of Integral Thinking

Holonic theory

Wilber introduced the concept of the *holon*. According to the holonic conception, the reality is not composed by wholes nor have parts, but is composed by unities whole/part or holons[53]. Any entity or unity of consciousness can be understood simultaneously as a whole or a part, any unity of consciousness, simple or complex, is a holon.

Holons create hierarchies, organizing reality by increasing profundity and transcendence. On the other hand, holons establish heterarchies of similar holons, defined by holonic equivalence[54]. Profundity and span are two complementary aspects of holons. Holons are forming a holarchy[55]. There are holons (whole/part) in the physiosphere, in the biosphere, and in the noosphere. Wilber proposed 20 tenets that the holons accomplish[56]. Among them, there are the dimensions of the holons. Each holon has four dimensions[57].

Agency. The tendency to be a whole. Aristotelian *entelechy*, morphic field (Sheldrake), canon (Koestler), relative autonomy and completeness, yang. It manifests the tendency towards self-preservation, autonomy, self-responsibility, self-esteem. It assumes in this sense fixed forms or patterns, among which there are the 20 tenets. Wilber named this dimension as 'deep structure'. In pathological forms, it's manifested as alienation and repression. It's the part of the holon that gives structure or is structuring. Oriented towards improvement and optimization, has an agent character.

Communion. The tendency to the relationship, to create bonds, participatory, to express their partnership, the ability to be part of a whole, attract other parties, relationship with something broader, *yin*, self-adaptation. This dimension can be called *agape* (evolutionary), in the sense of attracting other holons. Pathological forms: fusion and indissociation. Focused on the relationship with other holons, creating relationships of participation and inclusion.

Self-transcendence. Autotransformation, creative novelty, creativity[58], each holon is a new whole/part that has its own new forms of agency and communion. This is the

DOI: https://doi.org/10.30564/jiep.v2i1.603

impulse to experience freedom, to find cohesion and unity through a greater, deeper, and broader totality. Articulation by breaks of symmetry (Prigogine) not by a regrouping of the same. Evolution is the result of self-transcendence, also called by Wilber as *Eros*, that is, Spirit manifested in something else: matter, body, mind, soul, etc. In this dimension, the "*telos*" or purpose is originated. If self-transcendence is not achieved, "*Phobos*" (fear, regression, panic, contraction, and repression) is experienced. It is the part of the holon that goes beyond himself, his interior, source of novelty, ability to change. Transcendence assumes that any holon is under an evolutionary drive and presents novelty.

Self-dissolution/self-immanence. The morphogenetic gradient in the manifest field. This means not only a manifest reality with some kind of support in the manifested reality but also the potential to evolve. Preservation of the current level or regression to previous levels. Wilber conceptualized it as the instinct of death or *Thanatos*, a force opposed to *Eros*. Part of the holon that is focused on the preservation of itself, in the maintenance, preservation of the current state.

A representational schema with the four dimensions is:

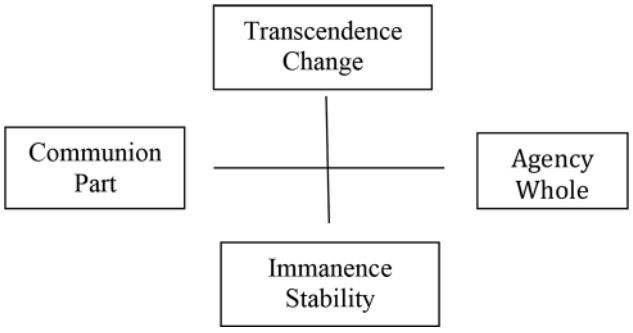


Figure 1. Holonic Dimensions

The holons are organized in a holarchy (hierarchies and heterarchies). Wilber characterized the global holarchy in the following holon: Immanence: objective world; transcendence: subjective states and stages of consciousness; agency: interobjective relationships and systems; communion: intersubjective shared cultural representations. By presenting the holonic theory we don't assume any kind of unnoticed dualism[59].

Conceptualization of Integral Thinking

Using the human understanding potential, integral thinking is the kind of thinking, which proceeds by following two steps:

Holistic consciousness

Identification, perception or recognition of an elevated holon addressed to understand the human consciousness or experience. The system can be: intellectual, philosophical, mystic-religious, psychological, related to self-knowledge or spiritual. Perceiving and understanding the transcendent dimension is a necessary step, otherwise, the holon wouldn't have the depth required and the system would collapse into a predictive system. Some relationship with a 'lived experience' is also necessary. It can be a perception, a dream, an intuition, a feeling, a sensation, a subjective thought, in short: a subjective holon. That experience has to be interpreted in the framework of the holonic system. In Jungian psychology, this recognition is explained as the operation of the transcendent function that unites consciousness and unconscious. Duality/polarity is a core element of integral theory[60]. That means some presence of collective unconscious materials in the consciousness, which can be denominated as the presence of the soul, the inner self, the presence of being or spirit, although there would be, of course, distinctions and clarifications to be made, because these aspects are not exactly the same. That union of opposites gives a sensation of security in the system, of internal truth, equivalent to the Aristotelian *nous*. This consciousness transcends all the previous levels, and there is a sensation of understanding them from the new position. Because transcendence is operating, there is, as a result, advance in profundity and consciousness. An elevation of consciousness is taking place. When this happens, because any holon is a whole, consciousness can be designated as holistic, or holonistic.

This level has been described as the change toward the second tier or the systemic level[61], the flexible flow level[62], and there is a correspondence with self-realization (Maslow), integral level (Gebser), fifth order of consciousness[63] and autonomous level (Loevinger, Cook-Greuter). It corresponds with vision logic inferior[64] and with the paradigmatic post-formal thinking stage. Commons and Richards[65] explained that at this stage a paradigm is created and their principles are applied to fields of different systems. Botella and Gallifa[66] studied the contextual character of postformal thinking and the difficulties to transfer schemas from one paradigm to another. That requires an additional effort.

Integral Thinking

Once the previous consciousness is developed for an elevated holonic structure, directed to explain human experience or consciousness, the next step is to advance understanding in depth the holonic traits and proprieties, the dimensions of the holon: immanence, transcendence, agency, communion. The thinking involved can be denominated as integral thinking. It's addressed to the recogni-

 DOI: https://doi.org/10.30564/jiep.v2i1.603

tion of the holonic structure dynamics. An important trait of integral thinking is to understand the whole-part relationships in the holarchy, going throughout the different holonic dimensions. Effectively, this second level requires going in depth to the relationship of the particular holon with other similar holons and with the entire holarchy, to understand in depth each dimension and the dynamics of that particular holon in space, time and movement. A consciousness act has the following dimensions: preservation, profundity, projective action and span[67]. The increase in these dimensions, especially span and profundity, results in an elevation of the consciousness. This elevation results in an emerging way to approach the experience.

This stage is equivalent to the global vision[68], the consciousness of constructs (Loevinger, Cook-Greuter), vision-logic superior o global mind interparadigmatic (Wilber), interparadigmatic (Commons and Richards). It's characterized by the integration of paradigms and to understand from here diverse complex systems. Gallifa and Botella[69] studied complexity as a contribution of differentiation and integration as independent dimensions. This level requires this double effort.

How Integral Thinking Works

An example of a holistic system in psychotherapy is bio-psycho-social-technical, being: bio (immanence), psycho (transcendence), social (communion) and technical (agency). Thinking integrally about that structure is relating each dimension with other holons and with the whole holarchy, finding the span of the model and profundity in different directions. For instance in what extent the Psychology (transcendence) under consideration is related to the transcendent dimension (the subjective quadrant) in the general holarchy. And the same can be done with the other dimensions. The internal actions implicated in these processes are what we defined as integral thinking. Transcendence is necessary.

The name of integral thinking means to find out the span and profundity of each dimension, making it communicable by reasoning. The first holistic level is more individual and contemplative oriented (consciousness), the second level requires the possibility of sharing it in a community using language (thinking). Integral thinking makes explicit what is implicit, by using the human capacity of understanding.

Although integral thinking is phenomenological in spirit, it's not enough to apply the phenomenological reduction[70], nor the eidetic. An implicit trait of integral thinking is that the eidetic process follows a general configuration. It utilizes as a guide the holonic structure, filling it with content, and using concurrently all the rest of modalities of thinking. The general holonic schema corresponds to a natural phenomenology of the consciousness[71].

The relevance of integral thinking is that allows thinking outside the schema "I'm right, you are wrong" characteristic of Modernity, and at the same time avoids the necessary reference to the authoritative sources of the premodern perspectives. Another asset is that goes beyond the relativistic thinking so characteristic of postmodern thought. Integral thinking doesn't substitute all of these modalities, but amplifies possibilities by thinking in a fourth way: the way of developing a thinking art (*tekhne*), complementary and integrating (excuse the repetition).

Integral thinking is an art, a *tekhne*. As an art, the production can be understood as an object outside the agent, and there is more than one possibility because it's oriented to *poiesis*, being a product of an *episteme* poietike, instead of an *episteme theoretike*, as is the case of the logical-argumentative thinking. Integral thinking, because achieves profundity in diverse dimensions and span, causes elevation of consciousness to a next level. Additionally by actualizing the *telos* contributes to the advance of human evolution.

The structure of integral thinking can be represented as a cross, circumscribed in a circle. This representation unites completeness (circle), meaning span, with four directions (cross), representing deepness in each direction. Is the simplest representation of a universal figure: The mandala. Integral thinking has a mandala-like structure. It's the thinking version of this universal internal-external form.

Finding the Holarchy in the Holon, Understanding the Holarchy from the Holon.

To understand and describe how the level of integral thinking (span and profundity in the four dimensions) can operate is particularly relevant. As it was proposed, the starting point is a particular identification of an elevated holonic system. Once identified, integral thinking proceeds to discover and describe the different holonic parts. Integral thinking involves going in depth in the different dimensions:

Transcendence. Is the necessary part that allows change and novelty. Is important to detect in any approach the *telos*, the actualization of new possibilities. That means to have a dynamics open to the new, to the future, that is not only a mechanical repetition of the past. For example, the presence of the subjective dimension can, at any moment, make evolve a complete system (consciousness, cultural, technical or scientific of any kind). Without transcendence, there is no possibility of an integral approach.

DOI: https://doi.org/10.30564/jiep.v2i1.603

When transcendence is reduced, the approach collapses into a predictive system.

Communion. This is the relationship with other holons. Is the sharing dimension, the presence of *agape*. Sometimes these relationships are not totally conscious and are manifested as worldviews, ways of understanding, language constructs. Is the world of implicit meanings, cultural rules, superego rules, limitation of human space or relationship with others.

Agency. Structure and unity of the holon and the potential to be projecting the internal structure to the exterior objective world. It can have the form of knowledge of the group. It can be a projection, transforming environment according to an objective, application of skills of any kind, competences, talent, knowledge of a group. It can be also an interpretative tradition shared by a group.

Immanence. Stability, persistence, habits, biologic conditionings, repetition of past schemas and patterns, the trend to preservation, that can include the possibility to preserve the good as in resilience.

Integral thinking allows finding the holarchy in a concrete holon and the other way round: to understand the whole holarchy from a particular holon. The holon more inclusive provides *telos* to the less inclusive. An example can be the Kabbalistic thinking that promotes the study of nested worlds, like Russian dolls, ones inside the others, all with the same characteristics. The same form can be found going from inside to outside, and vice versa.

The relations that integral thinking focuses on are:

Table 1. Relationship Holon-holarchy

To find:	A: Holon-holarchy	B: Holarchy-holon
1.In the part the part	The part of the holon has the same traits than the part of the holarchy	The part of the holarchy has the same traits than the part of the holon
2. In the part the whole	In the part of the holon there is the presence of the whole holarchy	In the part of the holarchy there is the presence of the whole holon
3. In the whole the part	The whole holon is a part of the holarchy	The whole holarchy is a part of the holon
4. In the whole the whole	The whole holon has the same qualities that the whole holarchy	The whole holarchy has the same qualities that the whole holon

Let's provide examples from different contexts in some of the cells. We could provide similar examples in the rest of the dimensions of the table:

1A. For example, the holistic psychotherapy presented. Saying bio (immanence) we have to apply the correspondent part of the whole holarchy (Wilber holarchy): Objective (immanence). That means objectivity of scientific findings. Integrating the findings of neuroscience to psychotherapy can be an example of this point.

3B. In the same example of the holon of holistic psychotherapy: Transcendence is psychology. In the psychological approach considered we have to find the whole holarchy. If a representation of the holarchy is the Wilber model, that would mean the characteristics of the psychology required, which needs to have the four dimensions: Object (I), Subject (T), Interobjective relationships (A), Intersubjective ones (C). The question for integral thinking can be: Does the approach consider scientific, subjective, cultural, and technical consciousness and evolution?

4A. Considering a medical specialty as a whole, and a specialist who knows in depth a system in the human body the thinker. Integral here means to apply the whole holarchy traits to the specialty. Knowing for example how the subjective, cultural, technical, scientific evolution traits can influence the whole particular biological system studied.

4B. An example can be found in the sentence: "A catcher of the rye" written as a title of the best seller of Salinger[72]. In this sentence, there is the presence of the whole message that the author wanted to communicate. The role of the book as a "catcher" of children by avoiding their fall in a precipice, a symbol for helping to maintain the infantile spirit in the described adult society characterized by the hypocrisy of customs and darkness in the real motives of the behaviors described in the book. A single sentence can contain all the condensed thinking of the author.

The purpose of integral thinking is to reflect on these relationships in order to have a more complete and balanced perspective, to evolve to a more elevated level of consciousness, and to develop an emergent way to deal with the particular experience.

This approach doesn't contradict the Aristotelian sentence: "the whole is greater than the sum of its parts". That principle is true in realities with minimal transcendence, but integral thinking, by considering transcendence, goes beyond this sentence without invalidating it. Integral thinking contemplates the relation part-whole and holon-holarchy in a broadened, more complex and complete way.

This approach explains well why syncretism is neither holistic nor integral. It can be, for instance, more truly 'integral' to belong to a tradition and going in depth to the holonic dimensions of the system, that merely adding perspectives without concern in transcendence and without span and profundity. The integral thinking goes beyond syncretic postmodern thought.

 DOI: https://doi.org/10.30564/jiep.v2i1.603

3. How Integral Thinking Works in A Thought System

Integral thinking can be applied to any problem, situation, art, science, or level of conscience, whether individual or collective, that involves human experience. As far as the application of integral thinking to a system of thought is concerned, first of all, the holonic schema has to be identified. Integral thinking results in an all-inclusive, transcendent, organizing, resilient art. That means that span and profundity should be considered. As any art (in the Aristotelian sense) more than one outcome is possible. The object (product of integral thinking) is an artifact (art-*factum*), the product of an art, which can be described as made by an integral art. Integral education produced integral educated individuals, integral medicine integral care and cure of patients, and so on.

How an integral system of thought can be identified? There are according to the development made four criteria for the system of thought considered:

1. Does the system have a holonic structure? All the dimensions have to be represented.

2. Is there in the system transcendence related to experience? There is the transcendent function. The system is holistic and not merely syncretic.

3. Are there relationships with other holonic systems? Is a proof of the character interparadigmatic. The system has 'span'.

4. Are there relationships with the whole holarchy through relationships whole-part? It's the specific trait of profundity in the different dimensions, a characteristic and necessary trait of integral thinking.

4. Integral Thinking and Educational Actuality

4.1 Holistic Education as A Potentiality

Gallifa[73] characterized holistic education from the traits of a holistic consciousness act. In this study, holistic education was characterized by the presence of an evidence-based program, technical-metacognitive projective action, the promotion of evolved cultural values, and an orientation to personal consciousness evolution. Holistic Education in this way defined is a *potentiality*, in Aristotelian terms, a field of pedagogical possibilities.

Nevertheless, the *actuality* of Education in the educational practice needs an adaptive kind of interaction to manage these cited dimensions of holistic education, using thinking skills in the practical arena. This *actuality* of the educational practice needs the special thinking skills that we characterized as integral thinking. Integral thinking is used to ponder these dimensions, balance them, sometimes deciding in seconds what decision to make in the practical context in order to integrate the four logics.

4.2 Integral Thinking in Educational Practice

Education is an *actuality* in space, time and movement. Holistic educational practice will benefit from the application of integral thinking because integral thinking is based in part-whole relationships that can be projected in the particular educational situation, composed by singular students, particular educational environment, and a holistic approach based in holistic consciousness. Using integral thinking between each one of the mentioned four dimensions of holistic education and the corresponding part of the whole holarchy (1A case of table 1), we'll obtain four logics:

Orientation to personal consciousness acquisition: Each human being has subjectivity and the possibility to develop diverse developmental lines as well as making evolve their self-system of consciousness. As has been explained, a particular system to understand the evolution of consciousness is necessary but, to a certain point, all systems share general principles. The levels of desire define the personal evolutional level[74]. The logic of sharing is more evolved and elevated that the logic of receiving for oneself alone. This dimension allows diversity in metaphysical, spiritual, religious, or philosophical options. Nevertheless, all of them share a wide common ground, around the common humanity of teachers and students. In addition, from the point of view of the student, the teacher is situated in the place of genuine sharing regardless of the system for understanding the evolution of consciousness is considered. Without this subjective dimension, education wouldn't have transcendence and therefore the approach wouldn't be holistic.

Evolution in values/culture: Each school or classroom is a human group and, as such, shares collective values. Diverse institutional logics are at work, institutional procedures, norms, and regulations. And all of them reflect implicit values. The evolutional logic here is that dialogic values and styles of social interaction are more evolved than authoritarian values. Sometimes these values are unconscious because they are expressed implicitly in routines and in unquestioned procedures. O the other hand the relationship of the school with the environment is also part of this logic. Schools can be part of networks, where there is also an exchange of giving and receiving. Democratic and dialogical systems of values are more evolved. Although democratic learning may be slower, this extra time is usually necessary to make possible the evolution in the group values.

DOI: https://doi.org/10.30564/jiep.v2i1.603

Evolution in learning of the knowledge shared by a group: This logic is about the acquisition of competences and techniques to develop a profession or to acquire more general competences useful for the practical life. This dimension manifests the presence of the societal needs and the logic of evolution of human life through productive capabilities[75]. Competence-oriented education is important and necessary. It's the 'learning by doing'. A wide range of abilities can be learned in this dimension. Even spiritual techniques can be understood as competence-based, like for example the case of meditation techniques.

Content acquisition: The purpose of any educational system is to guarantee content acquisition of sciences and humanities through specific curriculums. The evolution of knowledge is the evolutional logic of this point. Therefore the possibility of creating knowledge collaboratively is part of that dimension. Cooperative knowledge building is a more evolved way to acquire knowledge than knowledge transmission. Evidence-based programs have to be part of this educational systems logic.

For having an integral education practical approach, these logics have to be combined with the integral thinking modalities presented. Combining and balancing these logics in the educational *actuality* will cause that the holistic education approach considered, could have the potential to evolve, raise-above and situate the consciousness implicated in an emergent and elevated place. The consciousness implicated in modern schools and universities, as well as the consciousness of the postmodern ones, will be surpassed with that emergent model based in the use of integral thinking, which allows the elevation of consciousness in the educational *actuality* of a holistic education approach.

5. Integral Education

Education can be formal or informal, more planned or less. Each teacher or each student is at the same time a whole and a part (both can be recognized as holons). Integral thinking applied to education can be understood as a combination of all the parties (comprehensive education) and uniting them in a harmonic and balanced way (well-rounded).

To be an integral teacher is necessary to apply integral thinking to the educational practice. That means to be transforming these thinking processes into educational *actuality*. This means to be aware of all the dimensions taking part in a holistic education approach and to decide in the educational moments how to combine these dimensions. Especially important is not forgetting that education is about educating whole human beings. An integral teacher uses integral thinking to find the opportunity to amplify and elevate consciousness.

Integral education, in their more general meaning, can be understood as the emerging process of combining, by using integral thinking, the following future emergent trends in a comprehensive and well-rounded way. These trends are the previous four logics, now evaluated and projected towards the future, in inverted order:

Establishing a new relationship with knowledge. Educational systems, in Modernity and Enlightenment, focused on the transmission of knowledge from the empirical sciences. All the disciplines aspired to their scientific status and the model of specialization prevailed. The university also assumed the function to advance knowledge through scientific research. These roles, together with the extension of industrial societies, led to the teacher-student relationship becoming more vertical and transmissive. Today, however, technology facilitates access to knowledge as well as horizontality and the possibility of collaboration in the construction of knowledge. Educational systems will need to assimilate this potential. At the same time, for the acquisition of knowledge, it will be necessary to work on the basis of the evidence from educational research. With this evidence, the objective of reaching the diversity of all students and optimizing equity can be better approached. In this sense, the evolution of artificial intelligence, from which we are already beginning to see applications, will probably contribute to individualize and universalize learning and contribute to evaluating it more effectively.

Creating a closer relationship between education and life. The dynamism of society and the world outside of the schools and universities will have to cross even more the walls of the classroom. One of the essential dimensions must be a greater approach to the diverse and changing professions. This is the dimension of the master-apprentice relationship and must be fully incorporated. This implies taking into account and privileging the field of human will, which pushes the motivation to improve and contribute to the human environment. For that reason diverse approaches will have to be incorporated: the logic of the Arts, the perspective of the human action and work, the ability to propose solutions to real problems and improve them; all in order to better be directing the creativity and talent of students and to develop their skills and abilities. To facilitate this, guidance, coaching or personalized tutoring should be relevant, as well as better knowledge of the potential and singularity of the students to help to promote their vocations.

Paying more attention to the values ??of culture and organizations. The organizations where students learn

 DOI: https://doi.org/10.30564/jiep.v2i1.603

(schools or universities) are a model of society and culture, where students are learning collective values ??and acquire world-views. The culture of the school or the university prepares the culture for the societies of the future. Education is also training of future leaders and citizens. It will be necessary to promote schools or universities with more dialogical values, cooperative styles, and less vertical structures, which will be more evolved than the institutions we have inherited, too hierarchical and bureaucratic. This is the logic of the evolution of human consciousness for groups and organizations. Here the education of the future will also be able to incorporate the legacy of postmodern sensibility. There will be a need for an education with a truly more advanced consciousness to respond to the urgency and seriousness of the world's problems.

Giving a greater centrality to personal consciousness. It will be important also to recover the legacy of the pre-modern traditions that favored the Master-disciple relationship. This is the field of the evolution of subjective consciousness and the development of the inner dimensions. At this point the uniqueness of each person takes centrality. Personal happiness and fulfillment are at stake. The relationship or accompaniment is essential to achieve it. It is the awareness of the centrality of genuine authenticity, love, and esteem as a fundamental element of education. Without this dimension, education would remain as an excessive recreation of the past, without opening up the transcendence of each educational *actuality*. For that reason, we can talk about education and not just about teaching or strict instruction. To make possible this dimension we will have to use the common language of integral education to share the contribution to the development of consciousness coming from different traditions.

As has been argued, each educational event is composed, more or less consciously, for each one of these logics, which are also dimensions of human evolution. We propose for the future the integral education, which means a holistic educational approach where the previous dimensions have been taken into account in a conscious, balanced and coherent way.

6. Conclusive Thoughts

We described and defined integral thinking as a modality of thinking to deal with complexity and the different kind of relationships whole/part. Integral thinking is a structured form of thinking, based on the holonic structure and dynamics, which goes beyond the Aristotelian principle that 'the whole is greater than the sum of its parts'. It transcends it without negating their application in systems created by *epistemes*. Integral thinking operates defining the parts and the relation between them and the whole. Grounded in the holonic theory, is not merely holistic. It requires the consideration of the holonic structure but also to understand qualitatively the parts, and especially the relationship between other systems and parts/whole relations, including the relationships with the whole holarchy.

The adequate characterization of all the holonic dimensions is a distinctive trait of integral thinking. By going in depth in the different dimensions, transcendence included, integral thinking helps in increasing the profundity and to the elevation of consciousness. Pushes and pulls the advance of human evolution in different fields: Individual consciousness, culture, science or arts. Because it requires a high level of consciousness integral thinking is placed in an advanced location among the different modalities of thinking.

Postmodern thinking encourages integrated-syncretic approaches. Integral thinking, on the other hand, implicates span, but also conscious profundity. Integral thinking allows postmodern though to escape from syncretic relativism and helps Wilberian integral approaches to integrate the assets of premodernity, modernity, and postmodernity in one system of thought. The integral thinking core aim is to facilitate the evolution of human consciousness throughout the whole holarchy. In addition, integral thinking is useful to identify and/or to create integral perspectives in different fields, a contemporary true need.

Integral thinking can be applied to holistic education projecting part-whole relationships into the *actuality* of practical educational contexts. The result will be the evolution of the educational system considered to a more elevated level of consciousness. The emergent educative practice can be denominated as integral education.

In addition, integral thinking can help to the purpose of defining the dimensions to be considered in integral education practice and its agenda for the future.

References

[1] Aristotle. Metafisica (6a. Ed.) [Metaphysics]. Barce - lona: Iberia; 1984

[2] Baehr J. Sophia: Theoretical Wisdom and Contemporary Epistemology. In Virtues and their Vices. Eds. K Timpe & C Boyd. Oxford: Oxford University Press; 2014 p. 305.

[3] Cleary JJ. Nous as the Ground of Aristotle's Metaphysics? The Society for Ancient Greek Philosophy Newsletter. 164. [Internet]; 1993 [updated 2019 March 18 ; cited 2019 March 18]. Available from: https://orb.binghamton.edu/sagp/164. p 1

[4] Cleary JJ. Nous as the Ground of Aristotle's Meta - physics? The Society for Ancient Greek Philosophy Newsletter. 164. [Internet]; 1993 [updated 2019

  DOI: https://doi.org/10.30564/jiep.v2i1.603

March 18 ; cited 2019 March 18]. Available from: https://orb.binghamton.edu/sagp/164. p 1

[5] Baehr J. Sophia: Theoretical Wisdom and Contemporary Epistemology. In Virtues and their Vices. Eds. K Timpe & C Boyd. Oxford: Oxford University Press; 2014 p. 305.

[6] Trowbridge RH, & Ferrari M. Sophia and phronesis in psychology, philosophy, and traditional wisdom. Research in Human Development 2011; 8(2): 89-94.

[7] De Bono E. I'm right, you are wrong. New York: Vi - king; 1990

[8] Merleau Ponty M. Phénoménologie de la perception. [Phenomenology of perception]. París: Gallimard; 1945.

[9] Henriksen D., Richardson C., & Mehta R. Design thinking: A creative approach to educational problems of practice. Thinking Skills and Creativity 2017; 26: 140-153.

[10] De Bono E. Six thinking hats. London: Penguin; 1986.

[11] Gallifa, J. and Botella, L. The Structural Quadrants Method: A new approach to the Assessment of Construct System Complexity via Repertory Grid. Journal of Constructivist Psychology 2000; 13(1): 1-26.

[12] Wilber K. The theory of everything. Boston, MA: Shambhala; 2001.
Wilber K. Introduction to Integral Theory and Practice IOS basic and the AQAL map. AQAL Journal of Integral Theory and Practice 2005; 1(1): 1-38
Wilber K. The integral vision: A very short introduction. Boston, MA: Shambhala; 2007.

[13] Wilber K. The theory of everything. Boston, MA: Shambhala; 2001.
Esbjörn-Hargens S. Integral Education By Design: How Integral Theory Informs Teaching, Learning, and Curriculum in a Graduate Program. ReVision 2005; 28(3): 21-29
Helfrich PM. Ken Wilber's AQAL metatheory: An overview. Castaic, CA: Wildfire Media; 2008.

[14] Wilber K. Ken Wilber entrevistado por Raquel Torrent en 2016. Con preguntas de la Comunidad Integral Europea. [Internet]; 2016 [updated 2018 Dec 20; cited 2018 Dec 21]. Available from:https://asociacionintegral.wordpress.com/
Graves CW. Human Nature Prepares for a Momentous Leap. The Futurist 1974: 72-87

[15] Wilber K. Ken Wilber entrevistado por Raquel Torrent en 2016. Con preguntas de la Comunidad Integral Europea. [Internet]; 2016 [updated 2018 Dec 20; cited 2018 Dec 21]. Available from:https://asociacionintegral.wordpress.com/ p 10

[16] Wilber K. Ken Wilber entrevistado por Raquel Torrent en 2016. Con preguntas de la Comunidad Integral Europea. [Internet]; 2016 [updated 2018 Dec 20; cited 2018 Dec 21]. Available from:https://asociacionintegral.wordpress.com/ p 30

[17] Wilber K. Integral Spirituality: A startling new role for religion in the modern and postmodern world. Boston, MA: Shambhala; 2006.

[18] Wilber K. The theory of everything. Boston, MA: Shambhala; 2001.

[19] Piaget J. La psychologie de L'intelligence. [The psychology of intelligence]. París: Librerie Armand Colin; 1967.

[20] Campbell J. Los mitos: su impacto en el mundo actual. [The Myths: their impact in contemporary world] Barcelona: Editorial Kairós; 1994.

[21] Harari YN. Sapiens: A Brief History of Humankind. New York: Random House; 2014.

[22] Toulmin S., Rieke R. and Janik A. An introduction to reasoning. New York: Macmillan Publishing; 1984.

[23] Elqayam S. & Over DE. New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. Thinking & Reasoning 2013; 19(3-4): 249-265, DOI: 10.1080/13546783.2013.841591
Giménez-Camins C. and Gallifa J. Espinalt's Concept of Human Will and Character and Its Consequences for Moral Education. Ramon Llull Journal of Applied Ethics 2011; 2: 121-149
Evans JBT. Questions and challenges for the new psychology of reasoning. Thinking & Reasoning 2012; 18(1): 5-31, DOI: 10.1080/13546783.2011.637674

[24] Aristotle. Retórica. [Rhetoric]. Madrid: Alianza Editorial; 2014.

[25] Gardner H. Frames of mind: the theory of multiple intelligence. New York: Basic Books; 1983.

[26] De Bono E. Lateral thinking. A textbook of creativity. London: Pelican Books; 1977.
De Bono E. La revolución positiva. [Handbook for the positive revolution]. Barcelona: Editorial Paidós Ibérica, S.A.; 1994.

[27] Bereiter, C. and Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. De Corte, L. Verschaffel, N. Entwistle, and J. van Merriënboer (Eds.), Unravelling basic components and dimensions of powerful learning environments. EARLI Advances in Learning and Instruction Series [Internet]; 2003 [updated 2018 Dec 20; cited 2018 Dec 21]. Available from: http://ikit.org/fulltext/inresslearning.pdf

[28] Capurso V., Fabbro F. and Crescentini C. Mindful creativity: the influence of mindfulness meditation

 DOI: https://doi.org/10.30564/jiep.v2i1.603

on creative thinking. Frontiers in Psychology 2014; 4: 1020

[29] Jung CG. Analitical psychology: its theory and practice: the Tavistock lectures. New York: Vintage books; 1968. Jung CG. The basic writings of C.G. Jung. New York: The Modern Library; 1993.

[30] Gardner H. Frames of mind: the theory of multiple intelligence. New York: Basic Books; 1983.

[31] Sinnott JD. Post-Formal reasoning: The relativistic stage. In ML. Commons, FA. Richards and C. Armon (Eds.). Beyond formal operations: Late adolescent and adult cognitive development (pp 298-325). New York: Praeger; 1984.
Commons ML. and Richards FA. Four postformal stages In J. Demick (Ed.), Handbook of adult development. (pp 199-219). New York: Plenum Press; 2002.

[32] Botella L. and Gallifa J. A constructivist approach to the development of personal epistemic assumptions and worldviews. Journal of Constructivist Psychology 1995; 8(1): 1-18.

[33] Wilber K. Integral Spirituality: A startling new role for religion in the modern and postmodern world. Boston, MA: Shambhala; 2006.

[34] Commons ML. and Richards FA. Four postformal stages In J. Demick (Ed.), Handbook of adult development. (pp. 199-219). New York: Plenum Press; 2002.

[35] Usher M., Russo Z., Weyers M., Brauner R. and Zakay D. The Impact of the Mode of Thought in Complex Decisions: Intuitive Decisions are Better. Frontiers in Psychology 2011; 2: 37

[36] Ashlag Y. The Wisdom of Truth. 12 essays by the Holy Kabbalist Rav Yehuda Ashlag. Los Angeles CA: The Kabbalah Centre International; 2008.

[37] Gidley JM. Globally scanning for "Megatrends of the Mind": Potential futures of futures thinking. Futures 2010; 42(10): 1040

[38] Gidley JM. Globally scanning for "Megatrends of the Mind": Potential futures of futures thinking. Futures 2010; 42(10): 1046

[39] Wilber K. The theory of everything. Boston, MA: Shambhala; 2001.
Wilber K. Introduction to Integral Theory and Practice IOS basic and the AQAL map. AQAL Journal of Integral Theory and Practice 2005; 1(1): 1-38
Wilber K. The integral vision: A very short introduction. Boston, MA: Shambhala; 2007.

[40] Slaughter RA. What difference does 'integral'make?. Futures 2008; 40(2): 120-137.

[41] Choi I, Koo M & Choi JA Individual differences in analytic versus holistic thinking. Personality and Social Psychology Bulletin 2007; 33(5): 691-705.

[42] Upton J, Janeka I, & Ferraro N. The whole is more than the sum of its parts: Aristotle, metaphysical. Journal of Craniofacial Surgery 2014; 25(1): 59-63.

[43] Random House Unabridged Dictionary 2019

[44] Oxford Dictionary 2019.

[45] Oxford Dictionary 2019.

[46] McGregor SL. The promise of integral-informed FCS practice. Journal of Family and Consumer Sciences 2014; 106(1): 8

[47] Oxford Dictionary 2019.

[48] McGregor SL. The promise of integral-informed FCS practice. Journal of Family and Consumer Sciences 2014; 106(1): 8

[49] McGregor SL. The promise of integral-informed FCS practice. Journal of Family and Consumer Sciences 2014; 106(1): 9

[50] Gidley JM. Globally scanning for "Megatrends of the Mind": Potential futures of futures thinking. Futures 2010; 42(10): 1045

[51] Gallifa, J. and Botella, L. The Structural Quadrants Method: A new approach to the Assessment of Construct System Complexity via Repertory Grid. Journal of Constructivist Psychology 2000; 13(1): 1-26.

[52] Wilber K. The theory of everything. Boston, MA: Shambhala; 2001.
Wilber K. Introduction to Integral Theory and Practice IOS basic and the AQAL map. AQAL Journal of Integral Theory and Practice 2005; 1(1): 1-38
Wilber K. The integral vision: A very short introduction. Boston, MA: Shambhala; 2007.

[53] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 41

[54] Helfrich PM. Ken Wilber's AQAL metatheory: An overview. Castaic, CA: Wildfire Media; 2008.

[55] Koestler A. The ghost in the machine, New York, NY: MacMillan; 1967.
Helfrich, PM. Ken Wilber's AQAL metatheory: An overview. Castaic, CA: Wildfire Media; 2008.

[56] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. pp 43-85 and 527-532.
Gallifa J. Holonic Theory and Holistic Education. Journal of International Education and Practice 2018; 1(1): 36-46 DOI: https://doi.org/10.30564/jiep.v1i1.415

[57] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. pp 48-52 and 132-134.
Helfrich PM. Ken Wilber's AQAL metatheory: An overview. Castaic, CA: Wildfire Media; 2008.
Gallifa J. Holonic Theory and Holistic Education. Journal of International Education and Practice 2018; 1(1): 36-46 DOI: https://doi.org/10.30564/jiep.

DOI: https://doi.org/10.30564/jiep.v2i1.603

v1i1.415

[58] Whitehead AN. Process and Reality. New York: The Free Press; 1978.

[59] Gallifa J. Holonic Theory and Holistic Education. Journal of International Education and Practice 2018; 1(1): 36-46 DOI: https://doi.org/10.30564/jiep.v1i1.415

[60] Conte A. Polarities as a core element of integral theory. Journal of Integral Theory and Practice 2014; 9(2): 92-103.

[61] Graves CW. Human Nature Prepares for a Momentous Leap. The Futurist 1974; 72-87

[62] Beck DE. and Cowan C. Spiral Dynamics: Mastering Values, Leadership, and Change. New York: Wiley/Blackwell; 2005.

[63] Kegan R. In over our heads: The mental demands of modern life. Cambridge,
MA: Harvard University Press; 1994.

[64] Wilber K. Integral Spirituality: A startling new role for religion in the modern and postmodern world. Boston, MA: Shambhala; 2006.

[65] Commons ML. and Richards FA. Four postformal stages In J Demick (Ed.), Handbook of adult development. (pp. 199-219). New York: Plenum Press; 2002.

[66] Botella L. and Gallifa J. A constructivist approach to the development of personal epistemic assumptions and worldviews. Journal of Constructivist Psychology 1995; 8(1): 1-18.

[67] Gallifa J. Holonic Theory and Holistic Education. Journal of International Education and Practice 2018; 1(1): 36-46 DOI: https://doi.org/10.30564/jiep.v1i1.415

[68] Beck DE. and Cowan C. Spiral Dynamics: Mastering Values, Leadership, and Change. New York: Wiley/Blackwell; 2005.

[69] Gallifa, J. and Botella, L. The Structural Quadrants Method: A new approach to the Assessment of Construct System Complexity via Repertory Grid. Journal of Constructivist Psychology 2000; 13(1): 1-26.

[70] Merleau Ponty M. Phénoménologie de la perception. [Phenomenology of perception]. París: Gallimard; 1945.

[71] Wilber K. Eye to eye: The quest for a new paradigm. Boston, MA: Shambhala; 2001b.

[72] Salinger JD. The catcher in the rye. New York: Little, Brown & Company; 1951.

[73] Gallifa J. Holonic Theory and Holistic Education. Journal of International Education and Practice 2018; 1(1): 36-46 DOI: https://doi.org/10.30564/jiep.v1i1.415

[74] Berg Y. El poder de la Kabbalah [The power of Kabbalah]. Los Angeles CA: Centro de Kabbalah Internacional; 2006.

[75] Gallifa J. The Tekhne-logic revolution. Rethinking the 'interobjective' dimension of the Integral Theory. Consequences and relevance to Education. Creative Education 2018; 9(7): 1084-1104.

 DOI: https://doi.org/10.30564/jiep.v2i1.603

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep

ARTICLE

Challenging Excessive Fashion Consumption by Fostering Skill-Based Fashion Education

Madeleine Michell[1], Philip Henry*[2]

1, 2 School of Design, University of Leeds, United Kingdom

ARTICLE INFO

Article history
Received: 15th November 2018
Accepted: 13th March 2019
Published: 26th March 2019

Keywords:
Participatory design
Dressmaking
Sewing-Skills
Slow fashion
Co-design
Nudge theory
Sustainability

ABSTRACT

Academic arguments that fashion consumption habits need to change in order to support a more sustainable future are well established. This work explores a participatory design dress making methodology as a vehicle to trigger positive change in younger fashion consumers and help stem an attraction to fast fashion products. A key outcome of the study is that engagement in garment design and construction invoked an apparent positive emotional attachment not typically evident in fast fashion purchases. Participants demonstrated positive attitudes towards sustainability, alongside an awareness of the negative impact of fashion waste. However, at a relatively young age their evolving consumer habits show them to be drawn into the social drivers that are inherent in the idea of fast fashion. The perceived value of the work is the willingness of the students to participate. The apparent satisfaction they felt in their achievements and most significantly, their enthusiasm to continue with making. If this same enthusiasm could be captured in a new imagining of fashion and retailing the values of quality, longevity and individuality can remerge as the fashion ideal.

1. Introduction

The tensions between fast fashion and the conflicting desires for sustainability have been under scrutiny since the early 1990s. Despite many worthy initiatives, slow fashion and sustainable consumption have often been viewed as mere fanciful hopes. Cursory adoption policies and a limited commitment by the mainstream fashion world appear to have been the cause[1]. Equally problematic is the growing evidence in attempts to engage consumers in changing their buying habits. Amongst such a complex array of interrelated issues and priorities, people are left believing that the individual's impact is small and therefore pointless. This demonstrates how conflicting views on the subject matter can create confusion and hence counter-productivity. The objective of this work is to explore education and the development of traditional sewing skills as a means to offer school age consumers options as to how they choose to engage with the challenges of fast fashion consumerism.

Rather than attempting to challenge mind-sets already rooted in cycles of mass-consumption and fast fashion, this work looks to build on research exploring the potential to instigate a stronger personal attachment to garments by educating younger consumers. Ideally finding new

**Corresponding Author:*
Philip Henry,
Lecturer in Design Technology, School of Design, University of Leeds, United Kingdom;
Email: p.m.henry@leeds.ac.uk.

DOI: https://doi.org/10.30564/jiep.v2i1.403

ways to encourage a sense of irreplaceability, seldom identified within typical fast fashion purchases. One of the challenges circumnavigated in this research was an underlying need for dressmaking skills. In an earlier study Hirscher and Niinimäki[2] worked with 'pre-made' tunics offered to participants under the concept of 'half-made' to be customised by the individual. More recently Martin[3] addressed the same skills gap via the use of simplistic geometric shapes to create garments. The idea driving this new study, is that the use of a simple dressmaking technique, in conjunction with a participatory design culture, can help create a positive emotional attachment to the resultant garment; an experience that could potentially result in a longer life span of the garment, hence ideally reducing the desire to discard clothing and consume more.

2. Literature Review

2.1 Sustainable Fashion

Ideologies fostering sustainability were initiated during the 60s and continued to evolve through the 70s and 80s. The ideas then gained greater recognition in the 90s, post the perceived excesses of the 80s[4]. Regardless of any perceived negativity towards mass-consumption fashion brands, retailers continued to instigate and actively encourage consumerism by capitalising on the lower manufacturing cost afforded through globalisation strategies. Significantly for this study, Fletcher[5] identifies a distinct change in consumer attitudes during the economic growth of the 80s, leaving behind the values of the previous decade, quality, longevity and individuality. Throughout the last decade, consumer expectations are seen to have shifted to an ever-growing desire for cheap and convenient fashion, which is arguably becoming the ultimate obstacle to sustainability.

Nevertheless, the extent to which these shifts in consumer desire are the drivers for fast-fashion is difficult to know. The goal of manufacturing efficiency in itself is not new; 20th century economic theories such as 'Fordism' and 'Taylorism' are well recognised as catalysts of mass-production methodologies. Although these examples are born from the automotive and armaments sectors, the methods have been adopted across many industries. More prevalent in the fast fashion industry however, is that low costs are seemingly accepted as being synonymous with lower quality and the idea of 'functional obsolescence' is well established i.e. products with a deliberately short lifespan[6]. Perhaps, these are accepted characteristics of trend-driven clothing consumption, and a further highlight of the self-fulfilling 'fastness' of fast fashion, and the creation of desirable easy-to-consume, easy-to-replace products.

2.2 Slow Culture and Wellbeing

Counter to the values offered by the temporal satisfaction of consumerism and the culture of convenience, is the emerging idea of 'Slow', a culture that favours quality over quantity. Here a useful channel for parallel consideration is the food industry. When fast-food chain McDonald's opened a franchise in Rome, culinary writer Carlo Petrini began 'Slow Food' in protest against the rise of fast food culture. The ethos of the movement maintains: "A firm defence of quiet material pleasure is the only way to oppose the universal folly of Fast Life"[7]. The movement has gained a significant following across the globe and despite originating in the food industry, 'Slow' has since become a widespread concept defined by greater importance being placed on quality focused manufacturing processes, and investing in commodities that will withstand the test of time[5]. Minney[8] extends the ideals for slow fashion manufacturing by placing 'design, and therefore the designer, at the heart of the sustainable agenda, encouraging a 'mindfulness' approach across the whole supply chain.

A shift in the prioritisation of 'wellbeing' for consumers also offers some interesting parallels to consider. Again, linked to the food industry, the value of the comparison is the success achieved in changing purchasing habits from the pursuit of convenience to a positive change in lifestyle. An aim for change that is more about delivering a clear and direct value to the individual concerned, as opposed to a broader agenda more located in a sense of greater good. Statistics show that between 2012 and 2017 the 'Market Value of Health and Wellness' increased from just over 22 billion to nearly 25 billion Euros[9]. This indicates a shift from consuming convenient/fast and usually processed foods, to increased time spent exercising. The individual can directly feel the benefits of increased exercise and consumption of healthy food. For example, research suggests that time spent on personal health and wellbeing, results in higher self-value[10]. Further evidence that education can support a positive social impact can be seen in changing attitudes to tobacco and alcohol consumption[11][12]. Research shows significant reduction in younger age groups taking up smoking and alcohol particularly noted amongst sub-groups characterised by those with high physical activity, good mental health and in fulltime education or employment. A question driving this research is how to redirect the level of 'self-value' away from shopping for homogenised and expendable clothing, and to channel it towards a desire for individual expression realised through the experience of participatory design.

2.3 Fashion Consumption

 DOI: https://doi.org/10.30564/jiep.v2i1.403

The shopping experience, from one perspective is exciting and engulfed by feelings of personal satisfaction, deeming ethics and the environment a lesser issue when buying fashion[1]. Equally, clothing functions for individuals in that it 'protects' the human body from the elements[13]. The latter is a simple material need, however, the arguably more reflective, driven and non-material needs include more sophisticated concepts; for example identity, participation, recreation and freedom, are often materialised in a fashion form[5] In other words, the complex needs/motivation of individuals often depend on external factors (the consumption of products) to be satisfied[14].

While the constant proliferation of new fashion trends facilitates a seemingly endless progression of consumer aspiration, the positive feeling associated with such behaviours is evidenced to often be a short-lived experience. The work of Van Bowen[15] exposed the acquisition of material goods as leaving consumers feeling ultimately dissatisfied; any initial positivity is short lived generating a need for the feeling to be replaced with yet newer purchases. A self-perpetuating cycle, which has been described as one that oscillates between desire and disappointment, the initial love of a garment quickly transforms into discontent as its original appeal deteriorates[6]. A change to be addressed within this cycle is the anonymity of the consumer and their distance from the garments' manufacture that leaves an unbridgeable gap. Any potential receptiveness to the benefits of wearing a well made, and ethically sourced garment are not sufficiently tangible or real enough to encourage a genuine feeling of responsibility[16].

One of the reasons the external value of fashion items hold such personal appeal is social standing; the idea of being sufficiently skilled to make your own clothing, has some potential to be seen as aspirational. In the educational setting, the uptake of a skill is presented as an opportunity for further knowledge and offers a potential career pathway. Further, in an environment where such a skill is practised from the beginning of a school career, by the time a pupil leaves school they could be proficient enough to continue developing this skill. Therefore, the process of making, if practised regularly, has the potential staying power to create behavioural change[17]. Such thinking suggests that developing the skills of the consumer, to engage with the process of making/fashioning their own garments, has arguably some potential to challenge the cycle of consumerism[18].

2.4 Participatory Design

In this study, a participatory workshop is designed to encourage making skills specifically for an age group typically at an early stage of their fashion consumerism. The idea is to initiate the use of their skillset to fulfil their needs, as opposed to being entrenched in the habit of relying on a bought product. In such an environment, skillset development is presented as a longer lasting personal gain than the temporary gain felt after purchasing a product. Once an individual becomes proficient, the skillset can continuously be of benefit, and time spent engaging with the process can be recognised as worthwhile[17].

Beyond the skills value, there are other perceived benefits to be gained from engaging in participatory design; such experiences are understood to be a key component in unifying consumers with products. User participation creates a story that maintains a place in an individual's memory. The active participation in making a garment can establish a person-product relationship, therefore extending the life of the product[19]. Enjoyment and an accumulation of memories create a positive attachment to a product, increase the potential longevity, and indirectly promote more sustainable buying behaviour[20][21].

There are many arguments for the need to move the fashion industry towards a more sustainable future, as well as a necessity for our apparel buying choices to change. However, there is also a growing recognition that the aim must shift from trying to recondition current consumer habits. This approach is making little headway in the current retail environment, so there is consequently good reason to explore new approaches. Specifically, with an aim to refocus any satisfaction gained from the shopping experience, and instigate new methods of sustaining inherent human needs for creative interaction[22][1]. The process of active learning in dressmaking skills that Martin[3] proposes, echoes the 'Life of Action' expressed by Illich[23]. Individuals have the capacity to choose whether they buy items, or the freedom to make and use them. Here, active participation and interaction with personal power is perceived. The idea that developing dressmaking skills provides a more long-standing personal gain compared with the gain felt through the passive consumption of goods is also supported. Chapman[6] offers conclusions on how better design can produce product that people desire to keep for longer. Uncertainty still exists about the relationships between participation, increased emotional attachment and reduced consumption. However, there is no apparent evidence of previous studies employing a basic dressmaking method, in a paradigm that explores potential benefits for young teenagers. This work seeks to gain further insight into participatory design in a fashion context. Through the use of basic dress pattern shapes, it will explore self-sufficiency, self-worth, adventure and sustainability, using Martin's method as a vehicle to make clothing for all unique body types.

DOI: https://doi.org/10.30564/jiep.v2i1.403

2.5 Nudge Theory

Actively learning how to make our own clothing gives individuals more options and opportunity to better explore self-expression[3]. In defining the role of the designer, we establish that to design is to provide solutions in order to take us from a current situation onto the next one[24]. Arguably, knowledge of this aspect of design is fundamental for a sustainable future. "The Design Council reported that 80 per cent of a product's environmental impact is decided at the design stage"[25]. Waste can be argued as essentially a design flaw, making the education of the designer, maker, and user, regarding this challenge, a crucial step in encouraging change. In the context of this study, the important objective is to create positive reinforcement, and a sense of achievement in the acquisition of new skills. The aim is to present individuals with the means to make their own clothing, and in line with Nudge Theory[26], not radically change their habit by prohibiting them from buying garments. Instead the aim is to present an alternative way to create incremental changes in the drive towards a more sustainable future in fashion. Furthermore, consumption habits are known to relate to how individuals present their values, and how they want them to come to fruition[8]. The values of a commitment to developing dressmaking skills could instil dedication, patience, criticality and independence. Essentially, this study sets out to assess if knowledge of dressmaking skills has potential to encourage individuals to prioritise sustainable consumption.

3. Methodology

3.1 Study Design

The following research employs a multi-method approach, Methodological Triangulation (originally also referred to as 'mixed methodology') to obtain both quantitative and qualitative data through questionnaires and observation[27][28]. The design of the empirical dressmaking workshop based on the application of simple geometric pattern shapes combined Participatory Design methods to actively encourage an inclusive co-design environment, with Nudge Theory[26] being the driver for a change in the participants' thinking. The collaborative approach[29] allowed the collection of a new set of data on the potential value of inclusive design and positive enforcement through ethnographic observation. The objective of this research is to explore education, and the development of traditional sewing skills, as a means to offer school age consumers alternative options to engage with the challenges of fast fashion consumerism. The practice element of the methodology was delivered through a collaborative fashion skill workshop that took place over the course of two days.

The data collected, provides a platform to address the question of whether awareness of how to make one's own garments, can positively impact change in the way garments are used and cared for. Furthermore, the data assesses the ability of the participatory process to elongate the life of the clothing, reducing the necessity to buy more. An assessment of the dressmaking methodology presented by Martin [3] further aims to influence the field of creative education and move fashion towards a more sustainable future.

Questionnaires were strategically placed throughout the workshop experience, pre, post and six weeks following the workshop, allowing for a holistic evaluation of the results. The three questionnaires were designed to create a data set based on both qualitative and quantitative measure, by combining closed and open-ended questions to provide opportunity for a greater insight into participants' experience. Each set of questions was presented to participants at different stages of the workshop to evaluate their experiences in a managed way. Different types of closed questions were used including, dichotomy (yes/no), multiple choice, quantity and frequency of occurrence. Ranking questions were used to establish participants' attitudes towards the experience of the workshop and the likelihood of behavioural change. These included Likert scales, numeric scales, and behavioural scales[29]. The workshop and its strategic evaluation aim to disseminate knowledge of fashion practice.

3.2 Participants

The workshop, 'DRESS CODE: DIY' was located in the Art Department of UK based Secondary School for Girls, and comprised of 14 female participants between the ages of 13 and 15 (school years 9 and 10). In order to generate the desired immersive environment for the workshop an application process was used to select participants. Application letters where potential candidates expressing a high level of enthusiasm and evidence of some aptitude for active engagement in fashion practice were selected to take part.

3.3 Instruments

The participants employed a simple garment making process originated by Martin[3], making the evaluation suitably design process focused. Each participant was provided with the same set of instructions and necessary tools (a part-circle paper template, fabric, scissors, sewing machine, needles, tape measure and paper). All participants then used part circle-shaped pieces of fabric to construct their individual 'Segment Dresses'. To encourage individual engagement with creative choices, fabric colour selections were offered prior to the workshop. The 'shape'

 DOI: https://doi.org/10.30564/jiep.v2i1.403

method encourages basic understanding, through active participation. This simultaneously develops the practice skills required to make the garment.

Figure 1: Colour Selection and Guide Garment Example

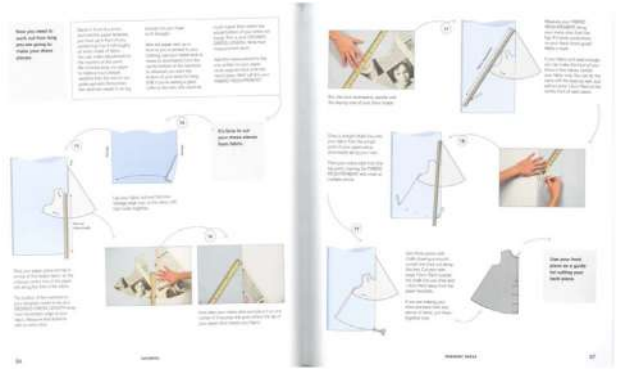


Figure 2: Set of instructions given to each participant reproduced from Martin 2016

The Workshop Domain

The researcher divided the participants into two groups of seven based on their initial progress in the task. For example, those who had created their patterns, and drawn them out onto the fabric, with the correct measurements, remained in Group One. Group Two spent the day working on their machine skills (setting up, seams, and zig zag finishes). On the second day Group One swapped with Group Two and in the afternoon, individuals continued working independently. Anecdotal observations and informal discussions were recorded throughout. This provided qualitative data, which served as a means to assess levels of engagement and the general behaviour of participants in the workshop environment.

The first written questionnaire was completed after an introductory presentation given by the researcher. The objective of the presentation was to give an overview of the problems facing the fashion industry. This included infographics on the negative impact of the current fashion industry and an overview of the researcher's industry experience. Current solutions, including upcycling and recycling were explained, along with the instructions for making the garment. Giving participants an overview of the expectations of the participatory workshop was essential to aid engagement from the outset.

The questions required respondents to provide information about their knowledge of fashion, their consumption habits, and their sewing skills. It was important to measure participants' interest in fashion prior to taking part in the workshop to provide a valid platform for comparing the data gathered at the conclusion of the workshop. At the end of the second day, participants completed questionnaire two, designed to evaluate engagement levels and collect participants' views on the workshop. The questions also aimed to explore whether participant involvement in the making of the garment would likely increase the participants' typical use of such a garment.

Figure 3: Collection of photographs showing the workshop in progress

The final questionnaire was delivered six weeks after the workshop, enquiring if the experience had positively impacted the relationship between participants and their 'Segment Dresses'. The impact of the workshop experience on buying behaviour was also measured in terms of whether participants had favoured new sewing projects over purchasing new garments.

DOI: https://doi.org/10.30564/jiep.v2i1.403

4. Results and Analysis

4.1 Observations

In addition to the three questionnaires, further data was collected through informal discussions and behavioral observations over the two-day workshop. It is interesting to consider at the outset of this discussion an element of contradiction that reflects the broader paradox. Despite their relatively young age, each of the participants expressed strong ethical credentials in their application letter, one defining herself as being an Eco-Prefect: "I am an Eco Prefect so am passionate about saving our environment and controlling the amount of waste that is produced". Consistently the selected group identified a strong interest in fashion design and the fashion industry, however they also evidenced knowledge of pertinent environmental concerns and issues surrounding waste: "As much as I love clothes, I feel that I do share the concern for the amount of fabric and materials that are wasted during the production of clothing". Another participant expressed concern writing that "waste is the last thing we want to hear when it comes to fashion as it is such a major part of our everyday lives, and because of that we need to find a way to use it sustainably". What is particularly interesting to this study is the participants' desire to improve on, or develop, the skills necessary to make their own clothes: "I would love the chance to work on my sewing skills as I do not get the chance outside of school". Further comments included: "I have been sewing since a very young age, but I know I could still learn a lot to make my work better" and, "I haven't ever had the chance to try sewing or working with fabrics, yet it is something that I would love to learn". Another participant wrote: "My ambition is to go to art school, I love to make new clothes out of my old favourites. I love a challenge because I think it is great to learn new things".

4.2 Questionnaire One

The initial questionnaire pre-workshop helps illustrate the nature of the broader challenge. When asked what they spent their money on, the majority selected clothes/fashion over other commodities and activities. These included, accessories, the cinema, eating out, books, iTunes and Apps. Although there was a general claim that garments were worn at least 10 or more times, and kept for a year and longer, the majority still claimed to purchase new clothing at least once a month. Despite any commitment to be an 'Eco-Prefect', the desire to be 'fashionable', and the drive to engage with the anticipated trends, proliferated via social media (36% of participants claimed social media had the strongest influence when buying a new outfit), is arguably a stronger lure, than any counter concern for the environment[30].

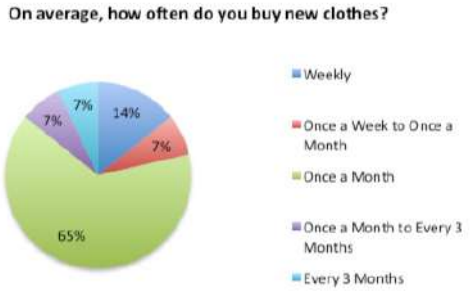


Figure 4: Pie Chart of participant responses to question 4 on the Pre-Workshop Questionnaire

As highlighted, the participants all identified a prior interest in dressmaking, and were keen to begin the workshop. However, it was apparent that they wanted to achieve the end result straight away. This became obvious through a reluctance to fully read and understand the instructions. Rather than rely on problem solving skills, participants requested to be directly told or shown what to do. This desire to be immediately satisfied was reminiscent of literature exploring factors that have contributed to individuals expecting such immediate solutions[31]. However, once participants began to take ownership, the distractions disappeared, as seeing their work evolve encouraged them to continue, indicating positive reinforcement. One individual who had made significant progress by the end of the workshop exclaimed: ''I just want to stay all night and finish it". Informal questions and empirical observations during the class further enforced the importance of fashion to the individual and the connotation relating to perceived social identity. At the same time, awareness of sustainability issues and the underlying concerns were commonplace amongst the participants. Their knowledge of these issues was mostly conveyed through education and exposure to conventional media.

Figure 5: Participants one word summaries of the workshop (responses to question one on the Post-Workshop Questionnaire)

 DOI: https://doi.org/10.30564/jiep.v2i1.403

4.3 Questionnaire Two

Questionnaire Two gathered information regarding the dressmaking experience of each participant. Overall the responses suggested that there was a general feeling of pride in what had been achieved, with 93% responding yes to the question of whether they were proud of what they had achieved. Participants selected their two most enjoyable things about the workshop from a list of six. The options were based on the structure and aim of the experience. The two most selected were gaining new sewing skills and being involved in a different creative activity. The four remaining options included, working with your peers, having two days out of timetable, making something for yourself, and building on previous experience. Out of all 14 participants, all except one felt excited by the option to make, alter and personalise garments after the workshop, and over half said they would consider making their next garment before purchasing it. A negative aspect of the workshop was expressed in relation to the expectation for more help and more time to complete the work. This may suggest a lack of independence, but perhaps measured against the desire to improve the quality of the resultant garments.

4.4 Questionnaire Three

In summary the final questionnaire distributed and collected six weeks post-workshop revealed that for the majority, an emotional attachment remained for their dresses. Of the participants, 21% had worn their dresses since taking part in the workshop and 64% of participants felt the experience of the workshop created an attachment to their dress. One participant wrote the reason being: "so much work went into it and it was the first time I made a garment with a sewing machine". Perhaps more interestingly, just over half of the participants had kept up their sewing skills and indeed gone on to independently start new sewing projects. For example, one individual wrote: "I am making a short / top combo to take to Greece", another explained: "I made a dress for my party". Further examples included, trousers, and a scarf. The remaining participants returned to their default purchasing habits, citing limited time and/or lack of confidence in their sewing skills. The majority of participants were unsure if the workshop experience had created fond memories in relation to their garment, and they were 50/50 on the workshop experience creating an emotional attachment to the dress. Overall, with 9 of the 14 participants agreeing it is unlikely they will discard the dress; involvement in the making process decreased the likelihood of discarding the garment.

Figure 6: Photograph of participants wearing their 'Segment Dresses' during the workshop

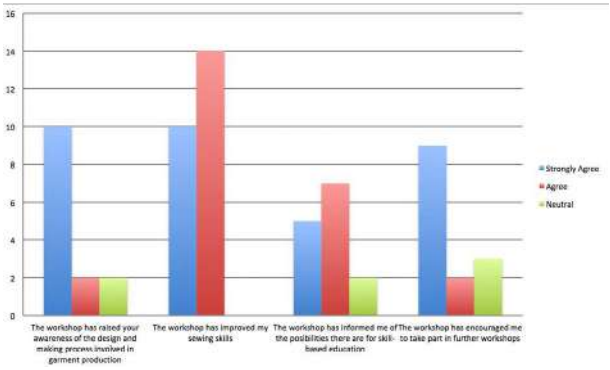


Figure 7: Bar Chart of participant responses to question two of the Post-Workshop Questionnaire

5. Discussion

The broader aim of the workshop was to offer young women an opportunity to develop a designer/maker skillset. A further aim was to moderate their fashion consumption, and ideally generate a greater emotional attachment to otherwise disposable garments. It is recognised that statistically the results are not significant, however, arguably they offer value in the potential to instigate change by empowering people to behave differently. To return to the analogy of attitudes to food and health discussed in the literature review; where people can identify a direct personal benefit clearly, they can be motivated to invest in change. The participants consistently recognised the negative impacts of fashion consumption on the environment, and equally how current shopping habits are questionable in terms of waste, verses sustainable impact. The underlying issues are widely taught and integrated into the standard curriculum, at the same time a current topic widely debated on standard media. Despite knowledge and understanding of the interrelated problems gained through education, without an alternative, they can be seen to simply respond in the same way as their older contemporaries.

The ideals surrounding slow fashion are deemed as fanciful without commitment from the mainstream fash-

DOI: https://doi.org/10.30564/jiep.v2i1.403

ion industry[1]. Arguably, consumers and their collective response are a key influence here. If there was a viable option, a participatory design methodology that better engaged consumers, and offered a level of personalisation through sewing skills, for example, there could be numerous mutual benefits. A key ingredient is the proliferation of sewing skills through mainstream education. However, there is the need for a creative response from the Retail sector to imagine new consumer experiences exploring co-design. Here digital technologies have potential to support mass-customisation from many platforms. Advancements in digital fabric printing, weaving, knitting and importantly pattern-cutting technologies, offer the means to evolve the desirable circular economies often discussed as an antidote to the global supply chain and the associated environmental issues. Engaging consumers in design and making, logically offers prospective empowerment with the means to be proactively involved in a solution.

A difficult to measure problem characteristic is the possibility of increasing the lifecycle of garments. Although the research presented offers no tangible evidence that the dresses will endure, Fletcher[5] identifies quality, individuality, and longevity as important values missing from the purchase decisions amongst current consumers. The pride and sense of achievement the participants collectively expressed at the close of the workshop are arguably triggers for an extended garment lifespan. In addition, it is not unreasonable to suggest that skills gained by the individual are equally triggers for the recognition of both making and design qualities in the garments they might buy. A greater first-hand appreciation of making skills could also underpin a greater sense of purchasing satisfaction, potentially countering the negative feelings and short-lived value identified by Van Bowen[15] and Chapman[6]. It is also reasonable to surmise that more discernible fashion choices, underpinned by new knowledge can perhaps support a changing in attitude, and help to counter the lack of responsibility consumers feel regarding the consequences of fast fashion as reported by the Business of Fashion[16].

6. Conclusion

This study was designed to determine the effect of teaching dressmaking skills as a means to instigate change in the buying habits of young women. Holistically, the findings suggest that the participatory design experience has the potential to nurture stronger relationships between young people and their clothing. Furthermore, the new knowledge and skills of dressmaking, gained through education, show a broader potential to encourage more considered fashion purchases. It is recognised that the aim to change consumer fashion buying habits is a complex problem, and attempts to recondition established patterns of consumption are making little headway. The tangible outcome for this study is that new behaviour can be fostered through an education experience helping to empower up and coming generations of consumers. An earlier and more widespread introduction of clothing design and making education is shown to have clear potential to contribute as a springboard towards a 'leaner economy'[32].

REFERENCES

[1] Sumner, M. 2017. It may not be possible to slow down fast fashion – so can the industry ever be sustainable? [Online]. [Accessed October 25, 2017]. Available from: http://www.independent.co.uk/life-style/fashion/it-may-not- be-possible-to-slow-down-fast-fashion-so-can-the-industry-ever-be- sustainable-a797003.html

[2] AL Hirscher, K Niinimäki (2013) Fashion activism through participatory design Proceedings of the Crafting the Future

[3] Martin, R. 2016. No Patterns Needed DIY Couture from Simple Shapes. London UK: Laurence King Publishing Ltd

[4] Feldman, E. 1992. Fashions of a decade: the 1990s. London: B.T. Batsford.

[5] Fletcher, K. 2014. Sustainable Fashion and Textiles Design Journeys (second ed.). Oxford: Routledge.

[6] Chapman, J. 2009. Design for (Emotional) Durability. Massachusetts Institute of Technology: Design Issues. 25(4), pp. 29-35

[7] Honoré, C. 2005. In Praise of Slow How a Worldwise Movement is Challenging the Cult of Speed. London UK: Orion Books Ltd.

[8] Minney, S. 2016. Slow Fashion aesthetics meets ethics. Oxford UK: New Internationalists Publications Ltd.

[9] Statistica. 2018. Market Value of Health and Wellness in the United Kingdom (UK) from 2012 to 2017 (in million euro0. [Online]. [Accessed February 27, 2018].

[10] Korn, L, Gonen, E, Shaked, Y, Golan, M. 2013. Health Perceptions, Self and Body Image, Physical Activity and Nutrition among Undergraduate Students in Israel. PloS ONE. [Online] 8(3) [January 15 2018]. Available from: https://doi.org/10.1371/journal.pone.0058543

[11] Office for National Statistics. 2017. Adult smoking habits in the UK: 2016 Cigarette smoking among adults including the proportion of people who smoke, their demographic breakdowns, changes over time, and e-cigarettes. [Online]. [Accessed December 2018]. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-018-5995-3.

 DOI: https://doi.org/10.30564/jiep.v2i1.403

[12] Ng Fat, L, Shelton, N and Cable, N. 2018. Investigating the growing trend of non-drinking among young people; analysis of repeated cross-sectional surveys in England 2005-2015. London , UK. BMC Public Health. [Online]. [Accessed October 2018]. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-018-5995-3

[13] Max-Neef, M. 1992 Development and Human Needs in: Elkins, P. ed(s). Real Life Economics: Understanding Wealth Creation. London UK: Routledge pp.197-214

[14] Thorpe, A. 2007. The Designer's Atlas to Sustainability. Washington: Island Press, p.118.

[15] Van Bowen, L. 2005. Experientialism, Materialism and the Pursuit of Happiness. Review of General Psychology. [Online] 9(2) pp.132-142 [January 14 2018]. Available from: http://psych.colorado.edu/~vanboven_rgp_2005.pdf

[16] Business of Fashion. 2017. Fashion Needs an Open-Source Sustainability Solution. [Online]. [Accessed October 13, 2017]. Available from: https://www.businessoffashion.com/articles/this-week-in-fashion/fashion- needs-an-ope n-source-sustainability-solution

[17] Gladwell, M. 2008. Outliers: The story of Success. United States: Little, Brown and Company.

[18] Fletcher, K. 2010. Slow Fashion: An Invitation for Systems Change. Fashion

Practice. 2(2), pp, 259-266

[19] Sung, K, Cooper, T and Kettley, S. 2015. An exploratory study on the links between individual upcycling, product attachment and product longevity. In: Cooper, T, Braithwaite, N, Moreno, M and Salvia, G (Joint Editors). PLATE Conference. 17-19 June. Nottingham, UK. Nottingham: Nottingham Trent University: CADBE, pp.349-356

[20] Schifferstein, H.N.J and Zwartkruis-Pelgrim, E.P.H. 2008. Consumer- product attachment: Measurement and design implications. International Journal of Design. 2(3) pp.1-13

[21] Oxborrow, L. 2015. Extending the Life of Clothes. Textiles. 42(3), pp.6-9 Fashion Revolution. 2017. What's Your Clothing Love Story? Loved Clothes Last. 1(2), pp.51-61

[22] Dickson, M, Cataldi, C and Grover, C. 2016. The Slow Fashion Movement. [Online]. [Accessed August 10, 2017] Available from: https://www.notjustalabel.com/editorial/slow-fashion-movement

[23] Illich, I. 1971. Deschooling Society. London: Calder and Boyers Ltd Martin, R. 2016.

[24] Friedman, K. 2016. Forward. In: Muratovski, G, Research for Designers a Guide to Methods and Practice (p.xxi). London: SAGE Publications Ltd.

[25] Hoskins, T. E. 2014. Stitched Up The Anti-Captialist Book of Fashion. London UK: Pluto Press.

[26] Thaler, R.H and Sunstein, C.R. 2009. Nudge: Improving Decisions About Health, Wealth, and Happniess. London, UK: Penguin Publishing Group.

[27] Yin, R.K. 1994. Case Study Research: Design and Methods. Thousand Oaks, CA: SAGE

[28] Bryman, A. 2006. Integrating Quatitative and Qualitative Research: How Is It Done? Qualitative Research. 6(1), pp.97-113

[29] Muratovski, G. 2016 Research For Designers A Guide to Methods and Practice. London UK: SAGE Publications Ltd

[30] Marsh, S. 2017. Girls suffer under pressure of online 'perfection', poll finds. [Online]. [Accessed October 27, 2017]. Available from: https://www.theguardian.com/society/2017/aug/23/girls-suffer-under- pressure-of-online-perfection-poll-finds

[31] Bobrow, E. 2014. In search of lost time. Why is everyone so busy? The Economist. [Online]. (Christmas Double Issue) [Accessed January 24, 2018]. Available from: https://www.economist.com/news/christmas- specials/21636612-time-poverty-problem-partly-perception-and-partly- distribution-why

[32] Flemming, D. 2016. Lean Logic: A Dictionary for the Future and How to Survive It. Vermont: Chelsea Green Publishing.

DOI: https://doi.org/10.30564/jiep.v2i1.403

BILINGUAL PUBLISHING CO.

Journal of International Education and Practice

http://ojs.bilpublishing.com/index.php/jiep

ARTICLE

The Comparative Effects of Cyclic Inquiry Model, Conceptual Change Text, and Traditional Instruction on Students' Understanding of Photosynthesis and Respiration in Plants

Salem Abedel Aziz Alkhawaldeh*

Faculty of Educational Sciences, Department of Curricula and Instructions, Al al-Bayt University, Mafraq, Jordan

ARTICLE INFO

Article history
Received: 21 November 2018
Accepted: 21 January 2019
Published: 29 March 2019

Keywords:
Cyclic inquiry model
Conceptual change text
Photosynthesis
Misconceptions
Science education

ABSTRACT

The aim of this study was to explore the effects of the cyclic inquiry model, conceptual change texts, and traditional instructions on promoting understanding of photosynthesis and respiration in plants. The data were obtained from 33 students in the first experimental group taught with cyclic inquiry model (CIM), 34 students in the second experimental group taught with conceptual change texts (CCT), and 34 students in the control group taught with traditional instruction (TI). After instruction, data were analyzed with analysis of covariance (ANCOVA) using pre-test scores and logical thinking scores as covariates. The results indicated the cyclic inquiry model (CIM) and conceptual change texts (CCT) treatment groups significantly outperformed the traditional instruction (TI) group in understanding the photosynthesis and respiration in plants. A statistically significant difference between two experimental groups was found in favor of the of cyclic inquiry model CIM.

1. Introduction

Research on students' understanding of scientific concepts in the past few decades has indicated that students hold many ideas that are different from those generally accepted by scientists. In the last two decades, there has been a number of studies that investigated students' misconceptions about photosynthesis and respiration in plants at middle and secondary schools[1–7].

For example, Haslam & Treagust[7] diagnosed understanding of photosynthesis and respiration in plants. The sample of the study consisted of 441 Australian students (grades 8-12). The results highlighted the consistency of the students' misconceptions across secondary year levels and indicated that a high percentage of secondary school students do not comprehend the nature and function of plant respiration and have little understanding of the relationship between photosynthesis and respiration in plants. Ozay & Oztas[6] studied the misconceptions held by 88 grade 9 students' (14-15 years old) in Turkey about photosynthesis and plant nutrition. Results revealed that students have conflicting, and often incorrect, ideas about photosynthesis, respiration and energy flowing. Svandova[3] investigated the common misconceptions of 108 lowest secondary school students (age 11-16 years)

**Corresponding Author:*
Salem Abedel Aziz Alkhawaldeh,
Professor, Faculty of Educational Sciences, Department of Curricula and Instructions, Al al-Bayt University, Mafraq, Jordan;
Email: skhawaldeh@aabu.edu.jo.

 DOI: https://doi.org/10.30564/ jiep.v2i1.335

in Czech Republic. The research showed that the students have many misconceptions about photosynthesis and respiration. Ameyaw[2] investigated 150 Senior High School (SHS) students' conception of photosynthesis and respiration in Ghana. The results gathered from the study showed that 31.5% of students did not know that glucose is the raw material for cellular respiration, and that water is produced as a by-product in aerobic respiration. Likewise, 23.6% and 29.9% of the respondents said Adenosine Tri- phosphates (ATPs) are not released at the end of aerobic respiration, and that anaerobic respiration does not occur in both plants and animals respectively. It also came to light that 36.7% of the respondent did not know that anaerobic respiration does not require oxygen for the reaction to proceed. Susanti[1] investigated the misconceptions of Biology education of Sriwijaya University in Indonesia about photosynthesis and respiration . The sample of the study consisted of 58 students'. The results indicated that: photosynthesis occurs continuously (37.9%), energy used for photosynthesis are light and heat energy (34.5%), plants take CO2 to respiration (47%), plants carry on respiration in the absence of light for photosynthesis (22.4%), respiration in plants occurs only in leaf cells (76.4%), and only animals that take O2 of photosynthesis to respiration (68.9%).

Alternative strategies have been offered by researchers to enhance meaningful learning in science. According to Novak[8], Conceptual change is necessity for meaningful learning to occur. The four conditions that are necessary for conceptual change to occur: (a) there must be dissatisfaction with existing conceptions, (b) the new conception must be intelligible, (c) the new conception must appear initially plausible, (d) the new conception must be fruitful[9]. One of the most successful techniques based on conceptual change approach is the use of conceptual change text[10][11]. In these texts, the students are asked explicitly to predict what would happen in a situation before being presented with information that demonstrates the inconsistency between misconception and the scientific conception. Several studies have reported the effectiveness of conceptual texts on creating conceptual change and promoting meaningful learning in students regarding many science[5][12-17].

Inquiry –based learning model has been also used extensively in science education to promote meaningful learning, beside the use of conceptual change, text prepared according to a conceptual change approach. The inquiry-based teaching approach is one of most successful approaches, and supported on knowledge about the learning process that has emerged from research[18][19]. In inquiry-based science teaching, students engaged in many of the activities and thinking processes that scientists use to produce new knowledge. Teachers encourages by science educators to replace traditional method instructional practices, such as emphasis on textbooks, lectures, and scientific facts, with inquiry based approaches that (a) engage student interest in science, (b) provide opportunities for students to use appropriate laboratory techniques to collect evidence, (c) require students to solve problems using logic and evidence, (d) encourage students to conduct further study to develop more elaborate explanations, and (e) emphasize the importance of writing scientific explanations on the basis of evidence[20]. Sandoval & Reiser[21] indicated that in order to build the inquiry-based classroom environment must construct a community of practice like the scientists work. The students take actions as scientists did, experiencing the process of knowing and the justification of knowledge, in authentic inquiry-based activities.

Currently, although studies in science education revealed the value of inquiry-based learning, teaching models of inquiry-based learning are diverse[22][23][24]. For instance, the inquiry cycle developed by Bruce and colleagues[25][26][27] consists of five stages: asking, investigating, creating, discussing, and reflecting (see Figure 1). Each stage in this inquiry cycle – seen as a process that provides learners with context-situated and content-specified learning experiences that help them explore the world in a connected fashion – can be embedded, interrelated or independent, depending on the situated learning needs. So, this cycle embraces an exploratory approach that motivates learners who have problems to be solved; engages learners through investigation, hands-on practice, collaboration, and dialogues; and stimulates learners' construction of meaning through the process of solving problems and then posing emergent questions.

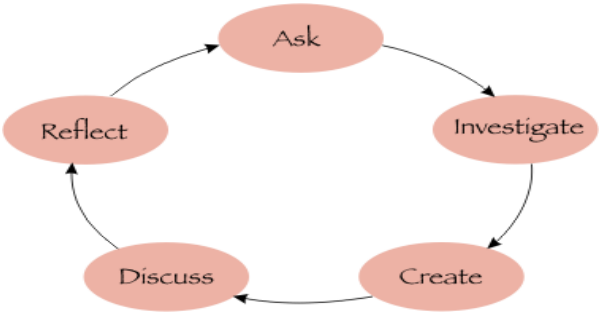


Figure 1. The inquiry cycle

The five stages in the process - ask, investigate, create, discuss, reflect-overlap, and not every category or step is present in any given inquiry. Each stage can be embedded in any of the others, and so on. In fact, the very nature of inquiry is that these phases are mutually reinforcing

DOI: https://doi.org/10.30564/ jiep.v2i1.335

and interrelated. Together, they comprise a cycle that can be used to inform and guide educational experiences for learners.

Ask

This stage, and the entire inquiry cycle begins with the desire to discover. Meaningful questions are inspired by genuine curiosity about real-world experiences and challenges[25]. The students raise and ask questions about the topics or issues, and then the teacher prepared a list of all questions related to the topics raised by students, and presents these questions to students to answer.

Investigate

Curiosity turns into action. Students gather information, study, design an experiment, observe, or interview. The student may recast the question, refine a series of query, or plunge down a new path that the original question did not, or could not, anticipate. The information gathering stage becomes a self-motivated process that is owned by the engaged student[28]. Investigating encourages students by to examine their topics using various sources of information and then plan out their creation and offers them an opportunity to navigate their inquiry[25].

Create

The student begins to make connections, as the information gathered in the investigation stage begins to coalesce. The ability to synthesize meaning at this stage is the creative spark that forms all new knowledge. The student now undertakes the creative task of shaping significant new thoughts, ideas, and theories outside of his/her prior experience[29]. On this step each group of students Writes a report includes all the knowledge , ideas, and information that have been discovered, and new conclusions reached that might be contribute to answering the main questions[30].

Discuss

Through discussion (or dialogue), construction of knowledge becomes a social enterprise; Students share their ideas and ask others about their own experience and inquires[28]. the discussion involves listening to the others and articulating their own understanding , helps them to achieve meaningful knowledge[25].

Reflect

Reflection is taking the time to look back at the question, the research path, and the conclusions made. The student steps back, takes inventory, makes observations, and possibly makes new decisions[29]. Has a solution been found? Do new questions come into light? What might those questions be?

Research has documented the effectiveness of cyclic inquiry model on enhancing meaningful learning in students regarding many science concepts[30-33]. For example Pansan & Nuangchalerm[32], compared learning achievement, science process skills and analytical thinking of fifth grade students who learned by using organization of project-based and inquiry-based learning activities. The sample used in the study consisted of 88 fifth grade students. Results revealed that the plans for organization of project-based and inquiry-based learning activities were appropriately efficient and effective. The students in both groups did not show different learning achievement, science process skills and analytical thinking. Albaaly[30] explored the effect of using the cyclic inquiry model in developing some of science processes and achievement in science among a sample consisted of 93 fifth grade students in Saudi Arabia. The results indicated that the cyclic inquiry model significantly outperformed the traditional treatment on the tests of science processes and the achievement in science. Abu al-Rukab[31] investigated the effect of cyclic inquiry model on the acquisition of scientific concepts and scientific thinking skills among a sample consisted of 147 fifth grade students in Jordan. Results indicated that there were statistical significant differences in acquisition of scientific concepts and scientific thinking skills attributed to the instructional model in favor of the cyclic inquiry model .

No research encountered in the literature that explores and compares the cyclic inquiry model instruction on conceptual understanding. Although the use of conceptual changes texts in science instructions are popular[12-17].

In an effort to promote conceptual understanding in science classroom, this research was conducted to examine the effects of the cyclic inquiry model (CIM), conceptual change texts (CCT), and traditional instruction (TI) on promoting understanding of photosynthesis and respiration in plants. It can be said that the main difference of the present study when compared to other studies is due to the cyclic inquiry model instruction factor on the conceptual understanding of photosynthesis and respiration in plants.

Statement of problem

This study was conducted to explore the effects of the cyclic inquiry model (CIM), conceptual change texts (CCT), and traditional instruction (TI) on promoting students' understanding of photosynthesis and respiration in plants. This topic is a fundamental part of biology curriculum and is considered abstract and difficult for students and teachers. Many researchers discussed the difficulties of teaching and learning photosynthesis and respiration in plant, others focused on students' conception related to photosynthesis and respiration. Less attention has been given to developing strategies or methods to eliminate these difficulties and remediate misconceptions, and improving photosynthesis and respiration in plants instruc-

 DOI: https://doi.org/10.30564/ jiep.v2i1.335

tion in basic biology classes. To improve understanding in basic biology classes, it is worthwhile to explore the effect of mode of instruction and cognitive variables on understanding of photosynthesis and respiration in plants. The main question is whether there are significant differences among the effects of CIM instruction, and CCT instruction, and TI instruction on students' understanding of photosynthesis and respiration in plants concept when photosynthesis and respiration in plants concept pre-test and TOLT scores are controlled as covariates.

2. Methodology

2.1 Sample

A total of 101 ninth- grade students, aged between 14 and 15 years (M=14.24, SD 0.42), enrolled in three classes in a basic-school in an urban area in Jordan. Each of the three classes instructed by the same Biology teacher (8 years of teaching experience) were randomly assigned as a CIM class (n=33), a CCT class (n=34), and a traditional class (n=34). Students in this study can be characterized as having middle-to high socioeconomic status (SES). Each class received identical syllabus-prescribed learning content. All the students fully participated in the study by attending classes, and completing the pretest and the posttest.

2.2 Instruments

The Photosynthesis and Respiration in Plants Concept Test

In this study the test developed by Haslam & Treagust[7] was used to determine students' conceptual understanding of photosynthesis and respiration in plants. It included a 13-item two- tier multiple choice test. The first tier of each item examined the content knowledge with two, three and four alternatives. The second tier consisted of reasons for the first tier, including a scientifically correct answer and three misconceptions[7]. A student's answer to an item was considered correct if the student answered both the content part and the reason part correctly. The test items were translated and adapted into Arabic. Content validity of each item was determined by a group of experts in science, science education, measurement and evaluation. The classroom teacher also analyzed the relatedness of the test items to the instructional objectives. The reliability coefficient computed by Cronbach's alpha estimates of internal consistency of this test was found to be 0.78, when both parts were analyzed. The test was administered to students in the three groups as a pretest, and post-test, to assess the students' conceptual understanding of photosynthesis and respiration in plant concepts over time.

The Test of Logical Thinking (TOLT)

In this study the test of logical thinking (TOLT), originally developed by Tobin & Capie[34], was used to determine the formal operational reasoning modes. The test was translated and adapted into Arabic by Abu Ruman[35]. It consists of 8 items designed to measure controlling variables, proportional, probabilistic, co relational, and combinational reasoning. The 8 items include two parts: an answer and a justification for the selected answer. The correct answer is the correct choice plus the correct justification. The internal consistency of the test was determined to be 0.82 using Conbach's alpha.

2.3 Treatment

This study was conducted over 3-week period in the first semester of the academic year 2017-2018. A total of 101 ninth -grade students were enrolled in three biology classes in a basic-school. Three classes, taught by the same biology teacher (8 years of teaching experience), were randomly assigned as a CIM class, a CCT class, and a traditional class. The classroom instruction for each group had two 45-minute periods per week. Students in all groups were exposed to same content for the same duration, and topic related to target concepts was covered as a part of the regular curriculum. Equal opportunities were considered to perform the activities in each group.

Students in the first experimental group were instructed with CIM instruction. Two separate CIM lessons, one for photosynthesis, one for respiration in plants, were designed by focusing on students' misconceptions and the objectives of the lesson. Lesson plans, including the objectives and detailed explanations of each phase of the CIM, were prepared as a guide. In the first phase (Ask) students' curiosity was prompted by asking questions about photosynthesis and respiration It is important that students formulate their own questions because they then explicitly express concepts related to photosynthesis and respiration. The second phase (Investigate) was designed to lead students to seek and create. Students or groups of students collect information, study, collect and exploit resources, experiment, look, and interview, draw….The third phase (Create) permitted students to merge collected information, they start making links. The ability to synthesize meaning is the spark which creates new knowledge. Students may generate new thoughts, ideas and theories that are not directly inspired. The fourth phase (Discuss) gave students the opportunity to share their ideas with each other, and ask others about their own experiences and investigations, they begin to understand the meaning of their investigations, comparing notes, discussing conclusions and sharing experiences. The final phase (Reflect) which requires taking time to look back, think again about initial question, the path taken, and the actual conclusion. Students look back and may take some new decisions.

DOI: https://doi.org/10.30564/ jiep.v2i1.335

Students in the second experimental group worked with CCT instruction method. Two conceptual change texts were prepared by the researcher considering four conditions proposed by Posner et al.,[9], dissatisfaction, intelligibility, plausibility, and fruitfulness. In each of the texts, students were introduced to questions and possible answers that may include misconception held by the student. Because of this technique, students were expected to be dissatisfied with their current conceptions. Then, scientifically accepted explanations that are more plausible and intelligible were described. Also, examples and figures were inserted into the texts for further help for students to comprehend the scientific concepts and realize the limitations of their own ideas.

Students in the control group were taught the topics of photosynthesis and respiration in plants by the teacher upon the basis of a lecture /discussion methods. The teaching strategy mainly relied on explanation by the teacher. Students read the topic from their textbooks in the classroom. Then, the teacher explained the concepts related to photosynthesis and respiration in plants by drawing examples on the board and illustrating important facts in the order as it appeared in the textbook. Specifically, the teacher used the chalkboard to write notes about the definitions of concepts, such as metabolism, enzyme, chloroplast, chlorophyll, mitochondria, ATP, cellular respiration, fermentation. After the teacher's explanation, concepts were discussed by the teacher via asking direct questions. The remaining time was taken up with the solving of various problems. The lesson ended with the students answering the questions orally. The main idea behind this teacher-centered instruction was to provide students with clear and detailed information. Students appeared to play a fairly passive role. Such instruction did not take students' misconceptions into account (see Appendix A).

3. Results

Descriptive statistics concerning the variables of the study were presented in Table 1. The table shows the means and standard deviations of the study variables. The mean scores on the TOLT, and the pretest reflected that students in each group had a medium level of formal-reasoning ability, and inadequate relevant prior knowledge, respectively.

Table 1 Means and Standard Deviations of the Variables

	CIM n=33		CCT n=34		TI	
Variable	M	SD	M	SD	M	SD
Pre-test	2. 54	0.94	3.06	1.41	3.65	1.07
TOLT	3.48	1.17	3.85	1.08	3.82	1.38
Post-test	7.03	2.35	6.76	2.61	5.85	1.52

Pre-test= *pre photosynthesis and respiration concept test, TOLT= test of logical thinking, Post-test= post photosynthesis and respiration concept test.*

To establish if there were significant differences in students' posttest means attributable to treatment, analysis of covariance (ANCOVA) was used to assess the effect of each treatment on students' understanding of Photosynthesis and respiration in plant concept test. ANCOVA was used in this study as it primarily served to adjust initial differences between groups attributed to the covariates. Table 2 summarizes the ANCOVA comparing the mean posttest scores of the performance of students in all groups.

Table 2 Summary of ANCOVA Comparing the Mean Posttest Scores of Students in the Three Groups

Source	*SS*	*df*	*F*	*p*	η^2
Pre-test	66.83	1	66.83	0.000	0.184
TOLT	45.519	1	45.519	0.000	0.133
Treatment	73.429	2	36.714	0.000	0.196
Error	297.3	96			

The analysis indicated significant effects for the covariates pretest score, $F(1, 96)=66.83$, $p=0.000$, and TOLT score, $F(1, 96)=45.519$, $p=0.000$. The results also revealed a significant treatment effect, $F(2, 96) = 36.714$, $p=0.000$ in favor of the experimental groups. Students in the experimental group who were engaged in the CIM instruction demonstrated better performance over the control group TI students ($p< 0.05$). Similarly, students who received CCT instruction scored significantly higher than students taught by traditional instruction TI with respect to understanding of photosynthesis and respiration in plant concepts ($p<0.05$), and students who received CIM instruction scored significantly higher than students taught by CCT instruction ($p<0.05$). Strength of the relationship between the treatment and posttest scores was strong $\eta^2=0.196$).

In order to determine which groups accounted for the difference found in the analysis of variance, a posterior comparison was made. The Bonferroni multiple comparisons was used to analyze paired contrast (Table 3). The photosynthesis and respiration in plants concept test adjusted mean scores of the two experimental groups engaged in cyclic inquiry model and conceptual change text instruction were significantly higher than the mean of the control group ($p<0.05$). The CIM group produced significantly higher adjusted mean score on the photosynthesis and respiration in plants concept test than did the CCT group.

Table 3 LSD Comparisons between Instruction Methods with Respect to Adjusted Mean Scores on the Photosynthesis and Respiration in Plants Concept Test

Comparisons*	Computed value	*p*
1 vs 2	0.88*	0.005
1 vs 3	2.23*	0.000
2 vs 3	1.35*	0.000

DOI: https://doi.org/10.30564/ jiep.v2i1.335

*p < 0.05. CIM 1: cyclic inquiry model. CCT 2: conceptual change text. TI 3: Traditional instruction.

When the proportion of correct responses and misconceptions determined by item analysis for the experimental and control groups was evaluated for both pre-posttests. Remarkable differences between the groups in favor of the experimental groups were found. The average percentage of students in cyclic inquiry model group holding the scientifically correct view had risen from 18.85% to 54.08%, a gain of 35.23%, the percentage of correct response of the students in conceptual change text group had increased from 23.54% to 52%, a gain of 23.46% after treatment. The percentage of correct responses of the students of the control group, however, increased from 28.08% to 45%, a gain of 16.92%. However, these results indicate low level of conceptual understanding even after the treatment.

Analysis of results indicated that students in experimental and control groups have many misconceptions about photosynthesis and respiration in plants. For instance, most students' conceptions were different from scientific meaning. Many students still had misconception" that oxygen gas is given out in largest amount by green plants in the presence of sunlight, because green plants only photosynthesize and do not respire in the presence of light energy". Research also, indicated that students believed that "green plants respire only at night, when there is no light energy because cells of green plants can photosynthesize during the day when there is light energy". In addition, several students had the misconception that " carbon dioxide is given off green plants in large amounts when there is no light energy at all because green plants respire only when there is no sunlight energy". Another misconception held by students was that " the most important benefit to green plants when they photosynthesize is the removal of the carbon dioxide from the air" . Moreover, several students thought that photosynthesis takes place in green plants in the presence of light and respiration takes place in green plants only when there is no light because green plants photosynthesize during the day and respire at night. A list of common misconceptions identified in the study were mentioned in Appendix B. Many of these misconceptions are typical misconceptions identified by other studies e.g. [7, 15, 16, 36].

4.Discussion

The purpose of this study is to investigate the effects of three types of instruction, the cyclic inquiry model (CIM), the conceptual change text instruction (CCT), and traditional instruction (TI), on 9th grade students' understanding of photosynthesis and respiration in plants. The photosynthesis and respiration in plants concept test developed by Haslam & Treagust[7] was administered to determine students' understanding of photosynthesis and respiration in plants. The test of Logical Thinking (TOLT) was used to determine the formal reasoning ability of students. The TOLT and pretest scores used as covariates in this study served mainly to reduce error variance. Results revealed that both covariates had significant effects on understanding of the target concept. This result is in agreement with previous studies results in the literature indicating that reasoning ability and prior knowledge have great influence on students' understanding of science concepts[37-40]. For example, Dogru-Atay & Tekkaya[37] reported that students' logical thinking ability accounted for a significant portion of variation in genetics achievement. BouJaude et al.,[38] pointed out that the main predictor of performance on conceptual understanding of chemistry was formal operational reasoning.

The results showed that the students in both cyclic inquiry model (CIM) and conceptual change texts (CCT) groups performed significantly better than students in the traditional instruction group with respect to photosynthesis and respiration concepts. The emphasis was given to students' misconceptions in both experimental groups. Students were engaged in activities that are intended to capture their attention, get them thinking about the subject matter, raise question in their minds, stimulate thinking and activate their prior knowledge. By these activities the evidence that students initial conceptions are insufficient and supported only partial understanding of the concepts were also provided.

What inquiry based learning (CIM) has in common is the active role of the students. Students are actively engaged in constructing knowledge. During each phase of the cyclic inquiry model, students are actively questioning and formulating problems, manipulating materials, observing and recording data, or analyzing data. By reflecting science processes or inquiry skills the cyclic inquiry model, allowed students to become active members in the process as they construct and understand scientific concepts. Because of the strength of cyclic inquiry model, students can see links among concepts and make connections between new learned concepts and existing concepts in their cognitive structures. Many interrelated facts and ideas are included in photosynthesis and respiration in plant concepts. Learners must relate the ideas and facts that form the concept to achieve meaningful conception of topics. In inquiry based learning the strategies used supported a change in students from passive to active learners. The activities involved in the cyclic inquiry model helped students recognize their prior conception, and helped them meaningfully learn through the connections among concepts and through developing reasoning skills.

DOI: https://doi.org/10.30564/ jiep.v2i1.335

This result is consistent with previous studies[30][31][2Similarly, results regarding the effectiveness of conceptual change text instruction (CCT) can be explained as follows: students activate and revise their prior knowledge and struggle with their misconceptions by involvement in activities. For example, students in the conceptual change text instruction became dissatisfied with their existing conceptions, which enabled them to accept better explanations to the problem that was introduced. In this way, students think about their prior knowledge and reflect on it. The conceptual change text definitely dealt with students misconceptions. It required students to construct an alternative schema to replace the misconception schema. The essential part of conceptual change text instruction was the social interaction provided by the teacher guided discussions. The students discussed the conceptual change texts with the teacher. The instruction encouraged intensive teacher-student interaction and student-student interaction, such a discussion environment allowed students to focus on learning, conceptual understanding, and mastering the task. However, students focused on concepts related to the subject that requires less conceptual restructuring students in the traditional Instruction group. This result is consistent with numerous previous studies investigating the effectiveness of the conceptual change text instruction[12-17].

The present study, However, has limitations to be considered. The study was conducted at a public school in an urban area by using whole classes. Data from other school districts and from different school types might give different findings. This study was limited to cell activities unit and 101 ninth graders. The results, therefore, may not be reliable if generalized beyond students enrolled in a similar situation.

5. Conclusion

The cyclic inquiry model instruction (CIM) and conceptual change text-oriented instruction (CCT) caused a significantly better acquisition of photosynthesis and respiration concepts and elimination of alternative conception than the traditional instruction (TI). This result supports previous research reporting that instructional strategies , which take into consideration the active role of the learner, and pre-existing concepts in the learners' cognitive structure, can promote better conceptual understanding. Additionally, this result supports the thought that it is not easy to eliminate misconceptions just by employing a teacher-centered and textbook-oriented instruction. It is necessary to eliminate misconceptions with the help of different methodologies that promote the active role of the learner rather than the traditional instruction, to create conceptual understanding and promote meaningful learning.

The findings of this study indicated that sound understanding could be reached with suitable instructional method. This study suggested the use of cyclic inquiry model and conceptual change oriented instruction as an alternative approach to traditional methods to remediate misconceptions.

References

[1] Susanti R. Misconceptions of Biology education student of teacher training and education of Sriwijiaya university to the concept of photosynthesis and respiration 2018; Journal of Physics: Conference Series 1022 012056.

[2] Ameyaw Y. Evaluating students' misconceptions of photosynthesis and respiration in Ghanaian senior high students. International Journal of Advanced Biological Research 2016; 6(2):202-209.

[3] Svadova K. Secondary school students' misconceptions about photosynthesis and plant respiration: Preliminary results. Eurasia Journal of Mathematics, Science & Technology Education 2014; 10(1): 59-67.

[4] Alkhawaldeh S, Al Olaimat, A. The Contribution of Conceptual Change Text Accompanied by Concept Mapping to Eleventh-Grade Students' Understanding of Cellular Respiration Concepts. Journal of Science Education and Technology 2010; 19: 115–125.

[5] Yenilmez A, Tekkaya C. Enhancing understanding of photosynthesis and respiration in plant through conceptual change approach. Journal of science Education & Technology 2006; 15(1): 81-87.

[6] Ozay E, Oztas F. Secondary students' interpretation of photosynthesis and plant nutrition. Journal of Biological Education 2003; 37(2): 68-70.

[7] Haslam F, Treagust, DF. Diagnosing secondary students' misconception of photosynthesis and respiration in plants using a two-tier multiple-choice instrument. Journal of Biological Education 1987; 21(3): 203-211.

[8] Novak J D. Meaningful Learning: The Essential Factor for Conceptual Change in Limited or Inappropriate Prepositional Hierarchies Leading to Empowerment of Learners. Science Education 2002; 86: 548–571.

[9] Posner MG, Strike KA, Hewson PW et al. Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. Science Education 1982; 66: 211–227.

[10] Mikkilä-Erdmann, M. Improving Conceptual Change concerning Photosynthesis through Text Design. Learning and Instruction 2001; 11: 241–257.

[11] Chambers S, Andre T. Gender, prior knowledge,

 DOI: https://doi.org/10.30564/ jiep.v2i1.335

interest, and experience in electricity and conceptual change text manipulations in learning about direct current. Journal of Research in Science Teaching 1997; 34(2): 107-123.

[12] Tas E, Gulen S, Oner Z et al. The effects of classic and web-designed conceptual change texts on the subject of water chemistry. International Journal of Elementary Education 2015; 7(2): 263-280.

[13] Sodervik I, Mikkila-Erdman M, Vilppu H. Promoting understanding of photosynthesis among elementary school student teachers through text design. Journal of Science Teacher Education 2014; 25(5): 581-600.

[14] Alkhawaldeh S. Prediction/discussion-based learning cycle versus conceptual change text: comparative effects on students' understanding of genetics. Research in Science and Technological Education 2013; 31 (2): 168-183.

[15] Balci S, Cakiroglu J, Tekkaya C. Engagement, exploration, explanation, extension and evaluation (5E) learning cycle and conceptual change text as learning tools. Biochemistry and Molecular Biology Education 2006; 34(6): 199-203.

[16] Alparslan C, Tekkaya C, Geban O. Using the conceptual change instruction to improve learning. Educational Research 2003; 37 (3): 133-137.

[17] Cakir O, Geban O , Yuruk, N. Effectiveness of conceptual change text oriented instruction on students' understanding of cellular respiration concepts. Biochemistry and Molecular Biology Education 2002; 30(4): 239-243.

[18] Abdi A. The effect of inquiry-based learning method on students' academic achievement in science course. Universal Journal of Educational Research 2014; 2(1): 37-41.

[19] Bransford JD, Brown AL, Cocking RR, Eds. How People Learn. Washington, D.C.: National Academy Press 2000.

[20] Secker CV. Effects of Inquiry-Based teacher practices on science excellence and equity. The Journal of Educational Research 2002; 95(3): 151-160.

[21] Sandoval WA, Reiser, BJ. Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. Science Education 2004; 88: 342-375.

[22] Moyer R, Hackett J, Everett S. Teaching science as investigations: modeling inquiry through learning cycle lessons. Columbus, OH: Prentice Hall 2006.

[23] Jarrett D. Inquiry strategies for science and mathematics learning. Portland, OR: Northwest Regional Educational Laboratory 1997.

[24] Short K, Laird J, Ferguson, and M, et al. learning together through inquiry: from Columbus to integrated curriculum. York, ME: Stenhouse 1996.

[25] Bruce B, Lin C. Voices of youth: podcasting as a means for inquiry –based community engagement. E-Learning 2009; 6(2): 230-240.

[26] Bruce B, Bishop A. Using the web to support inquiry-based literacy development, Journal of Adolescent and Adult Literacy 2002, 45(8): 706-714.

[27] Bruce B, Davidson J. An inquiry model for literacy across the curriculum. Journal of Curriculum Studies 1996, 28(3): 281-300. http://dx.doi.org/10.1080/0022027980280303

[28] Casey L, Bruce BC, Martin, A, et al. (2009). Digital Literacy: New Approaches to Participation and Inquiry Learning to Foster Literacy Skills among Primary School Children. Dublin: Centre for Research and Innovation in Learning and Teaching, National College of Ireland. Available from http://hdl.handle.net/2142/9765.

[29] Bruce B, Jakobsson E, Thakkar U, et al. Democratizing Bioinformatics research in high school classroom. American Educational Research Association (AERA) Annual Meeting, Chicago IL, 2003; April 21-25.

[30] Albaaly A. The effectiveness of using cyclic inquiry model (CIM) in developing some science processes and the achievement in science among fifth grade pupils at the elementary stage. International Journal for research in Education 2012; 31: 259-284.

[31] Abu al-Rukab A. The effect of cyclic inquiry model on the acquisition of scientific concepts and scientific thinking skills among fifth grade students in Jordan. Unpublished Master Thesis, Al –Bayt University, Mafraq, Jordan 2015.

[32] Panasan M, Nuangchalerm P. Learning Outcomes of Project - Based and Inquiry - Based Learning Activities. Journal of Social Sciences 2010; 6 (2): 252-255.

[33] van Rens L, van der Schee J, Pilot A. Teaching Molecular Diffusion Using an Inquiry Approach: Diffusion Activities in A Secondary School Inquiry-Learning Community. Journal of Chemical Education 2009; 86(12): 1437 – 1441

[34] Tobin K G, Capie W . The development and validation of a group test of logical thinking. Educational Psychology Measurement 1981; 41: 413– 423.

[35] Abu Ruman K. The relationship between formal thinking and science process skills and achievement in science of first scientific secondary grade students. Unpublished MS Thesis, University of Jordan, Amman, Jordan 1991.

[36] Sanders M. Erroneous ideas about respiration: the teacher factor. Journal of Research in Science Teach-

  DOI: https://doi.org/10.30564/ jiep.v2i1.335

ing 1993; 30: 919–934.
[37] Dogru-Atay P, Tekkya, C. Promoting students' learning in genetics with the learning cycle. The Journal of Experimental Education 2008; 76: 259-280.
[38] BouJaude S, Salloum S, Abd-El-Khalick, F. Relationships between selective cognitive variables and students' ability to solve chemistry problems. International Journal of Science Education 2004; 26: 63–84.
[39] Kang S, Scharmann L C, Noh T. Reexamining the role of cognitive conflict in science concept learning. Research in Science Education 2004; 34: 71–96.
[40] Johnson MA, Lawson AE. What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? Journal of Research in Science Teaching 1998; 35: 89–103.

Appendix A
Comparison of Instructional Methods

Cyclic Inquiry Model Instruction	
Ask	The teacher introduced concepts and issues about the subject (photosynthesis for example) of the lesson, and encourage students to ask themselves about concepts and issues they find important and engage in conversation about these issues with their peers.
Investigate	Curiosity turn into Action. Students carry out experiments and inquiry activities, and practice different science processes skills in order to gather knowledge and information to answering questions in the previous stage.
Create	Students construct meaning by engaging in hands-on learning activities, in other words, students learn to transform the thought developed earlier through asking and investigation into substantive creation of meaning.
Discuss	Students share their ideas and ask others about their own experience and inquires. The discussion involves listening to the others and articulating their own understanding , helps them to achieve meaningful knowledge.
Reflect	The teachers gives students time to look back at initial questions, the research path and the conclusions made, and then invoke further questions derived from the reflection of current experience.

Conceptual Change Text Instruction
The texts were given to the students before the instruction.
The teacher directed the students to read it before the class hour and bring it to the class.
Students were informed about the new instruction, the nature of the conceptual change text, and how they would use it during the instruction.
Students read a paragraph in which a question was posed to arouse students' interest in the subject and to analyze their pre-conceptions.
Students shared their ideas about the answer with the class. The teacher did not intervene and did not give any feedback during this process.
Answers that are not scientifically correct (misconceptions) about the concept that were provided in the text were read aloud by one of the students.
Students were asked to compare their conceptions with these misconceptions.
The scientifically correct explanations of the concept that are intelligible and plausible were provided to guide students in considering why the misconceptions could be wrong.
The teacher asked whether anything related with the explanation surprised the students to help the students reconstruct the concepts.
Images, figures, and pictures were used to help students visualize the concepts while reading the text.
In addition, the history of science, such as cell activities, equations of photosynthesis, cellular respiration, and fermentation.

 DOI: https://doi.org/10.30564/ jiep.v2i1.335

Traditional Instruction

Teaching strategy relied on teacher's explanation. The teacher used the chalkboard to write notes about the definitions of the concepts, such as; cell activities, photosynthesis, enzyme, cellular respiration, fermentation, and write equations related with photosynthesis and cellular respiration.
After the teacher's explanation, concepts were discussed by teacher-directed questions.
The focus of the instruction was on problems related with cell activities, photosynthesis, and cellular respiration.
No experiments or hands-on activities were performed by the students related with the topics.

Appendix B

Common misconceptions identified by Photosynthesis and Respiration in plants Concept Test

Misconceptions
Carbon dioxide is used in respiration which only occurs in green plants when there is no light energy to photosynthesize
Respiration in plants takes place in the cells of the leaves only
Respiration is the exchange of carbon dioxide and oxygen gasses through plant stomata
Green plants take in carbon dioxide and give off oxygen when they respire
Green plants e do not respire in the presence of light energy, they respire only at night when there is no light energy.
Green plants respire only during daylight because green plants do not respire, they only photosynthesize
Photosynthesis provides energy for plant growth
Plants respire when they cannot obtain enough energy from photosynthesis and animals respire continuously because they cannot photosynthesize

DOI: https://doi.org/10.30564/ jiep.v2i1.335

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep

ARTICLE

Investigating the Psychometric Impact of Negative Worded Items in Reading Comprehension Passages with a 3PL Cross-Classified Testlet Model

Yong Luo[1*] Junhui Liu[2]

1. National Center for Assessment, Riyadh, Saudi Arabia.
2. University of Maryland, College Park, Maryland, USA.

ARTICLE INFO

Article history
Received: 19 December 2018
Accepted: 11 March 2019
Published: 30 March 2019

Keywords:
Negative wording
Bifactor model
Cross-classified testlet model
Validation

ABSTRACT

Negative worded (NW) items used in psychological instruments have been studied with the bifactor model to investigate whether the NW items form a secondary factor due to negative wording orthogonal to the measured latent construct, a validation procedure which checks whether NW items form a source of construct irrelevant variance (CIV) and hence constitute a validity threat. In the context of educational testing, however, no such validation attempts have been made. In this study, we studied the psychometric impact of NW items in an English proficiency reading comprehension test using a modeling approach similar to the bifactor model, namely the three-parameter logistic cross-classified testlet response theory (3PL CCTRT) model, to account for both guessing and possible local item dependence due to passage effect in the data set. The findings indicate that modeling the NW items with a separate factor leads to noticeable improvement in model fit, and the factor variance is marginal but nonzero. However, item and ability parameter estimates are highly similar between the 3PL CCTRT model and other models that do not model the NW items. It is concluded that the NW items introduce CIV into the data, but its magnitude is too small to change item and person ability parameter estimates to an extent of practical significance.

1. Introduction

Negatively worded (NW) items are often recommended to be included along with positively worded (PW) ones in psychological inventories to address acquiescence (e.g., Kieruj & Moors, 2013).[34] A common practice is to reversely code NW items and treat them the same as PW ones. Such a practice, however, depends heavily upon the assumption that there is no wording effect associated with the NW format used in NW items, which is usually not the case. Studies (e.g., Chessa & Holleman, 2007)[12] have shown that the cognitive process involved in answering NW items is different than that in dealing with PW items, and consequently NW items display different psychometric properties such as lower item-total score correlation (e.g., Roszkowski & Soven, 2010).[49] Consequently, some

**Corresponding Author:*
Yong Luo,
Senior Psychometrician, National Center for Assessment, Riyadh, Saudi Arabia;
Email: jackyluoyong@gmail.com.

 DOI: https://doi.org/10.30564/ jiep.v2i1.418

researchers (e.g., van Sonderen, Sanderman, & Coyne, 2013; [56] Zhang, Noor, & Savalei, 2016) [62] caution against the use of NW items, arguing that such items introduce extraneous variance and hence pose a threat to construct validity.

Factor analysis techniques have been routinely applied to psychological inventories containing NW items to investigate the wording effect. A two-correlated-factor model is often used (Deemer & Minke, 1999;[14] Gitchel, Roessler, &Tuner, 2011;[21] Magazine, Williams, & Williams, 1996;[42] Roszkowski & Soven, 2010)[49] to model the PW and the NW items, assuming that one factor representing the negative wording effect and the other the positive wording effect. The bifactor model has also been used (e.g., Lindwall, Barkoukis, Grano, Lucidi, & Raudsepp, 2012),[37] with the general factor hypothesized to measure the latent construct of interest, and the two secondary factors represented the positive and negative wording effects. Wang, Chen, and Jin (2015)[58] decisively pointed out the logical flaw of treating the positive wording effect as a separate secondary factor in the previous bifactor modeling approach and used a different bifactor model in which only the negative wording effect is modeled as a separate secondary factor.

In contrast to the psychological measurement literature where a consensus regarding the use of NW items is lacking, in educational testing community use of the NW format is usually cautioned against (Haladyna, 2004;[22] Haladyna & Downing, 1989a;[23] 1989b;[24] Haladyna, Downing, & Rodriguez, 2002).[26] Note that the NW items in educational testing often involve adding at the stem or option level a negative word such as not or except, while in psychological measurement they can either add negative words or use vocabulary that is opposite to the measured construct (e.g., use of the word sad in a scale measuring happiness). It is further recommended that in cases where it is necessary to use in an educational test the NW format (adding a negative word), the negative word "…should be stressed or emphasized by placing it in bold type, capitalizing it, or underlining it, or all of these" (Haladyna, 2004, p. 111).[22] One of the reasons for such recommendations against the NW items is that educational tests are usually high-stakes and students are too motivated to allow acquiescence bias to materialize. Another reason, as will be discussed later, is that the addition of a negative word in educational test items tends to change item psychometric properties.

Despite the suggestion that NW items should be avoided in high-stakes testing, they are still occasionally used in some educational tests, although per Haladyna's advice (e.g., 2002;[26] 2004),[22] negative words in those items are often emphasized. Research on NW items in educational measurement often utilizes experimental studies to investigate how NW items perform in contrast to their PW counterparts, and to date, there have been no studies applying a factor analysis approach to model the effect of the NW format in educational tests.

In this study we investigate the wording effect of NW items in a high-stakes English proficiency reading comprehension test using a model similar to the bifactor model used by Wang, Chen, and Jin (2015).[58] Our model is similar to that of Wang et al. (2015)[58] in that we also model the negative wording effect as a secondary factor independent of the primary factor, which is English proficiency in this case. The similarity notwithstanding, there are several major differences between the two models. First, guessing is expected to exist in our data due to the use of MC format and as a result, we include a pseudo-guessing parameter in our model to account for guessing. Since the mathematical equivalence between the item response theory (IRT) and factor analysis (e.g., Takane & de Leeuw, 1987;[52] Kamata & Bauer, 2008) [33] does not extend to models with pseudo-guessing parameters, our model is not a factor analysis model per se. Second, instead of using a general testlet model (Li, Bolt, & Fu, 2006)[35] that is the IRT analog of a bifactor model, we use a testlet model (Wainer, Bradlow, & Wang, 2007)[57] that is a constrained version of the general testlet model (e.g., DeMars, 2006;[15] Rijmen, 2010)[47] to model the negative wording effect. We believe the testlet model is more appropriate in the current study due to its estimation of testlet variance, which allows for a straightforward interpretation of the magnitude of variance caused by the wording effect. As will be discussed later, this variance is irrelevant of the primary latent construct of interest and its magnitude indicates how much of a threat it is to the test validity. Last, since our data were drawn from a reading comprehension test, the passage effect may cause items within the same passage to be locally dependent. For NW items within passages, we hypothesize they exhibit dual local dependence due to the passage and wording effects and in order to simultaneously model both effects, a cross-classified testlet model is warranted. Consequently, we use a three-parameter logistic (3PL) cross-classified testlet response theory (CCTRT) model to answer the following research questions:

Does the NW format introduce construct irrelevant variance (CIV) into the test?

If yes, what is its magnitude?

How does failure to model such CIV affect item and ability parameter estimates?

The remainder of this paper is organized into four

  DOI: https://doi.org/10.30564/ jiep.v2i1.418

sections. We start with a review of relevant studies in the educational measurement literature that investigate the wording effect of NW items. In the second section, we introduce the 3PL CCTRT model, followed by analysis of the current data in the third section. In the last section we end our paper with discussions and conclusions.

2. Studies Regarding NW Items in Educational Testing

Most studies focusing on the psychometric effect of NW items in the context of educational measurement adopt experimental designs in which the performances of two versions of a small number of items, one with NW format and the other PW format, were compared. For example, Terranova (1969)[55] found that the NW items were more difficult than their PW counterparts, and test reliability in his experiment was not affected by the NW format. Similarly, Dudycha and Carpenter (1973) [18] found that NW stems increased the item difficulty, whereas the item discrimination was not affected. In another study, Cassels and Johnstone (1984)[10] found that simply changing item stems from the NW format to the PW format with the options remaining constant lowered the item difficulty considerably, and they attributed such changes of item difficulty to the additional thinking stage required by NW format. Similar findings were also presented by Caldwell and Pate (2013)[8] that stem negation increased item difficulty. Johnstone (1983, p. 115)[31] indicated that the reason for increased item difficulty due to the NW format is that "…ideas in a negative form occupy twice as much space in the working memory as positive forms". Similarly, Abedi (2006)[1] listed the NW format as one of the linguistic features that might affect comprehension.

While the aforementioned studies consistently find that the NW format is associated with increased item difficulty, there are other studies that say otherwise. For example, Tamir (1991; 1993)[53][54] found that for items requiring low cognitive reasoning, the NW format did not affect the item difficulty; it was only when combined with requirements for high cognitive reasoning that the NW format increased item difficulty. Downing, Dawson-Saunders, Case, and Powell (1991),[17] and Rachor and Gray (1996)[44] found that the NW format had no effects upon item difficulty. In another study (Harasym, Price, Brant, Violato, & Lorscheider, 1992),[27] it was found that the NW format lowered both item difficulty and test reliability, and the researchers attributed the decrease of item difficulty to that the NW format inadvertently provided cues to the correct answer.

Other than focusing on how NW items behave in contrast to their PW counterparts, Casler (1983)[9] took a different approach by investigating whether emphasizing the negative words in the NW items alleviated the psychometric effect caused by the NW format. He found that with the negative words underlined, the NW items became easier to the high ability students and harder to the low ability ones, and the item discrimination power increased; if the negative words were capitalized, the item difficulty decreased while the item discrimination remained unchanged.

Apparently, the majority of abovementioned studies find that the NW format changes the psychometric properties of an item. Consequently, it has become a widely-accepted notion that the NW format is undesirable and should be avoided in the context of educational measurement. As stated by Haladyna (2004, p. 117):[22]

Avoid Negative Words Such as Not or Except. We should phrase stems positively, and the same advice applies to options. The use of negatives such as not and except should also be avoided in options as well as the stem. Occasionally, the use of these words in an item stem is unavoidable. In these circumstances, we should boldface, capitalize, italicize, or underline these words so that the test taker will not mistake the intent of the item.

To date, most studies regarding the psychometric effect of the NW format in educational testing literature utilized experimental designs to investigate how item properties such as item difficulty and discrimination change, and only one (Casler, 1983)[9] compared the psychometric effects caused by NW items with and without emphasizing the negative words within. There are no studies that use a bifactor model approach similar to that adopted by Wang, Chen, and Jin (2015)[58] to explore whether the NW items form a separate factor due to their NW format. We argue that such a NW factor, if existent, constitutes a source of CIV that can be a major threat to test validity (e.g., Hahadyna & Downing, 2004).[25] Messick described the role that CIV plays in educational testing as the following:

The major point here is that educational achievement tests, at best, reflect not only the psychological constructs of knowledge and skills that are intended to be measured, but invariably a number of contaminants. These adulterating influences include a variety of other psychological and situational factors that technically constitute either construct-irrelevant test difficulty or construct-irrelevant contamination in score interpretation. (Messick, 1989, p. 216)[43]

Despite the lack of consensus on how the NW format changes item difficulty, the majority of studies, as discussed previously, indicate that the NW format changes item difficulty due to the introduction of CIV. Similarly, Downing (2005, pp. 141-142)[16] stated that

 DOI: https://doi.org/10.30564/ jiep.v2i1.418

"the additional test difficulty introduced into the measure by poorly crafted and flawed item formats is an example of construct-irrelevant variance." Following his line of reasoning, we hypothesize that the NW format in items is a source of construct-irrelevant variance (CIV) and therefore, it is important to find out what its magnitude is and whether it affects model parameter estimates to an extent of practical significance.

It should be noted that since the data used in this study were drawn from a reading comprehension test, passage effect may constitute another source of CIV. In the next section we will discuss that although our interest is not in the passage effect per se, failure to model the passage effect might result in inaccurate estimation of the wording effect and consequently, we model both sources of CIV simultaneously with a CCTRT model.

3. The Testlet Model and Its Cross-Classified Extension

One pivotal assumption of item response theory (IRT) is local item independence, which can be expressed using the following equation (Reckase, 2009)[45]

$$P(\mathbf{U}=\mathbf{u}|\theta)=P(u_1|\theta)P(u_2|\theta)...P(u_I|\theta), \quad (1)$$

where $\mathbf{u}$ is a response vector to a test with I items, $P(\mathbf{U}=\mathbf{u}|\theta)$ is the probability of obtaining the response vector $\mathbf{u}$ for an examinee whose latent ability is θ, and $P(u_i|\theta)$ is the probability of obtaining a score u_i. Equation 1 states that after conditioning on the latent ability, the response to any item in the test is statistically independent of that to another item. In other words, an examinee's latent ability should be the only force that drives his or her item responses and, if there is another factor that affects the item response, this independence assumption is violated and local item dependence (LID) occurs. As listed by Yen (1993),[60] in real testing situations LID can occur due to various factors such as speededness, item or response format, and passage dependence. Numerous studies (e.g., Ackerman, 1987;[2] Chen & Thissen, 1997;[11] Zhang, 2010)[61] have shown that LID can result in biased estimation of item parameters, overestimation of test reliability, premature termination of computer adaptive testing, errors in equating, and erroneous classifications of examinees.

Due to the serious psychometric consequences that can be caused by LID, various methods (e.g., Bradlow, Wang, & Wainer, 1999[6]; Braeken, Tuerlinckx, & De Boeck, 2007[7]; Hoskens & De Boeck, 1997;[28] Rosenbaum, 1988[48]) have been proposed to address the issue of LID. Among them, a popular approach that has been applied extensively to address LID is the testlet response theory (TRT; Bradow, et al., 1999;[6] Wainer, Bradlow, & Wang, 2007),[57] which models LID among items within the same cluster by introducing a random effect parameter denoting the person specific testlet effect. The probability of answering item j correctly by person i in a 3PL TRT model is given as

$$p_i(\theta_j) = c_j + \frac{1 - c_j}{1 + e^{-a_j(\theta_i - b_j + \gamma_{id(j)})}}, \quad (2)$$

where θ_i is person i's latent ability, $\gamma_{id(j)}$ is person i's latent ability on testlet d, and a_j, b_j, and c_j are the discrimination, difficulty, and pseudo-guessing parameters of item j. For $\gamma_{id(j)}$, its variance $\sigma^2_{\gamma_{jd(i)}}$ indicates the magnitude of LID among items within the same testlet. As can be seen from equation 2, TRT only allows the modeling of one source of LID. When it is suspected that dual LID may exist, TRTM seems inadequate due to their incapability of handling more than one source of LID simultaneously.

It is not uncommon for test data to display dual LID due to the existence of two item clustering factors. For example, in PISA assessment items based on the same scenario may fall into different content categories: here scenario and content are two item clustering factors that may cause dual LID. Another example is that in language testing, items within the same listening comprehension passage may have different item formats, thus making format and passage two possible sources of LID. In the current study, dual LID is also suspected to exist due to two item clustering effects, namely the passage effect and the wording effect due to the NW format. To address the issue of dual LID, Jiao, Wang, Wan, and Lu (2009)[29] proposed a 3PL CCTRT model, which is an extension of the 3PL TRT model, to address dual LID in scenario-based science assessment items. Its equation is given as

$$p_i(\theta_j) = c_j + \frac{1 - c_j}{1 + e^{-a_j(\theta_i - b_j + \gamma_{id_1(j)} + \gamma_{id_2(j)})}}, \quad (3)$$

where $\gamma_{id_1(j)}$ is person i's latent ability on testlet d_1 caused by the first source of LID, $\gamma_{id_2(j)}$ on testlet d_2 caused by the second source of LID, and the other terms remain the same as in equation 2. As can be seen, if either source of LID has a variance of zero, equation 3 reduces to equation 2 and the 3PL CCTRT model becomes the familiar 3PL TRT model.

The difference between a TRT model and a CCTRT model can be visualized with diagrams. Assuming that we have a reading comprehension test with two passages (each has three items) and two of the six items (Items 2 and 4) have the NW format, Figure 1 provides a visual presentation of the TRT and CCTRT models that can be used to model difference sources of LID. As can be seen, the TRT model in the left panel only models LID due to the passage effect, and its CCTRT counterpart models dual LID due to both the passage and NW effects. One commonality between these two models is that there are

DOI: https://doi.org/10.30564/ jiep.v2i1.418

no arrows or curves linking the factors, which indicates that the factors are orthogonal to each other.

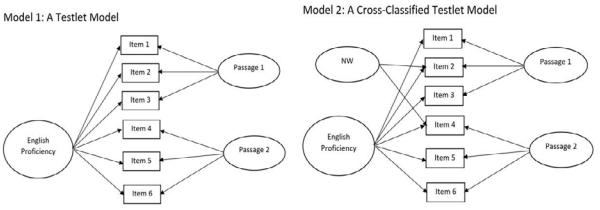


Figure 1. The traditional testlet model and the cross-classified testlet model

In addition to the 3PL CCTRT model, the Rasch version of CCTRT models has also received some attention in literature. Xie (2014)[59] proposed a cross-classified Rasch testlet model in the hierarchical generalized linear model (HGLM; Kamata 2001)[32] and investigated the consequences of failing to model the dual LID properly. She found that ignoring either source of LID leads to inaccurate estimation for item difficulty, ability parameter, and testlet effect parameters. Similar to the relation between the Rasch model and the 3PL model, Xie's model is a special case of the 3PL CCTRT model in equation 3 with the guessing parameter c_j constrained to zero and the discrimination parameter a_j to one across items. Jiao, Kamata, and Xie (2015)[30] extended Xie's model to its multilevel case and showed that the multilevel cross-classified Rasch testlet model can be accurately estimated using Markov Chain Monte Carlo (MCMC) methods implemented in OpenBUGS.

4. Methods

4.1 Data

The data were drawn from item responses of 1,839 students who took a high-stakes English proficiency test used for admission and placement purposes in the Middle East. For the purpose of this study, we focused on the reading comprehension section which consists of 40 items.

Table 1 lists descriptive statistics (p value and item total correlation) and other relevant information of these 40 items. The column named p provides the classical test theory (CTT) based item difficulty index, based on which we observe that the test is slightly difficult to the examinees since most p values are below 0.50. The column named r_{pb} provides the item total correlation values, most of which are between 0.3 and 0.5 with some lower than 0.2 (e.g., RC17).

The column named Item Type tells whether an item is a discrete one or belongs to a particular passage, and the column named NW tells whether an item is a NW one. As can be seen, 37 out of the 40 items are nested within reading comprehension passages and among them, seven items use NW format. Among these seven items, two use a negative word at the stem level and the other five at the option level, and all the negative words are bolded and capitalized for highlight purposes. Per our previous discussion, we hypothesize that these seven items display dual LID due to the passage effect and the negative wording effect.

Table 1 Item Statistics and Relevant Information

Item	p	r_{pb}	Item Type	NW	Item	p	r_{pb}	Item Type	NW
RC1	0.39	0.25	Passage 1		RC21	0.35	0.24	Passage 5	
RC2	0.50	0.44	Passage 1	Yes	RC22	0.46	0.42	Passage 5	
RC3	0.30	0.20	Passage 1		RC23	0.41	0.37	Passage 5	Yes
RC4	0.43	0.43	Passage 1		RC24	0.46	0.46	Passage 5	
RC5	0.43	0.49	Passage 1		RC25	0.43	0.45	Passage 5	
RC6	0.36	0.27	Passage 2		RC26	0.32	0.44	Passage 5	
RC7	0.33	0.33	Passage 2		RC27	0.36	0.29	Passage 5	
RC8	0.33	0.17	Passage 2	Yes	RC28	0.42	0.54	Passage 5	Yes
RC9	0.35	0.30	Passage 2		RC29	0.26	0.18	Passage 5	Yes
RC10	0.32	0.35	Passage 2		RC30	0.45	0.50	Discrete	
RC11	0.36	0.44	Passage 3		RC31	0.34	0.31	Passage 6	
RC12	0.46	0.46	Passage 3		RC32	0.35	0.37	Passage 6	
RC13	0.48	0.46	Passage 3		RC33	0.37	0.45	Passage 6	
RC14	0.45	0.43	Passage 3		RC34	0.23	0.30	Passage 6	Yes
RC15	0.27	0.32	Passage 4		RC35	0.31	0.36	Passage 6	
RC16	0.24	0.06	Passage 4		RC36	0.29	0.17	Passage 7	
RC17	0.17	0.09	Passage 4		RC37	0.32	0.19	Passage 7	
RC18	0.35	0.21	Passage 4		RC38	0.27	0.20	Passage 7	
RC19	0.35	0.39	Passage 4	Yes	RC39	0.47	0.48	Discrete	
RC20	0.36	0.39	Discrete		RC40	0.36	0.25	Discrete	

4.2 Analytic Procedure

Before we model such dual LID with a CCTRT model, it is necessary to determine which dichotomous IRT model should serve as the base model. Correspondingly, we estimate the 1PL, 2PL, and 3PL models with OpenBUGS and use Akaike's information criterion (AIC; Akaike, 1973),[3] Bayesian information criterion (BIC; Schwarz, 1978),[50] and the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002)[51] to determine the best fitting model, which is used as the base model in the subsequent analyses. It should be noted that AIC and BIC were originally developed in the maximum likelihood estimation (MLE) framework, and here we use their Bayesian analogues that are computed with the

 DOI: https://doi.org/10.30564/ jiep.v2i1.418

posterior mean of deviance as described by Congdon (2003).[13]

After the base model is determined, we build a corresponding TRT model in which the item clustering effect due to passage dependence is modeled, and a CCTRT model in which the two item clustering effects due to both the passage dependence and the NW format are modeled simultaneously. This particular CCTRT model is treated as the true model in the current study. Similarly, we estimate the TRT and CCTRT models with OpenBUGS and compare model fit using AIC, BIC, and DIC. We also compare the item and ability parameters between different models to investigate whether failure to model any item clustering effect translates into practical significance in terms of parameter estimate differences.

4.3 Estimation

All three models were estimated via MCMC algorithm implemented in OpenBUGS. It should be noted that Stan, an emerging Bayesian software program, may be a better choice for estimating complex IRT models due to its sampling efficiency (e.g., Luo & Jiao, 2018;[40] Luo & Liang, 2019).[41] OpenBUGS was chosen in this study due to its convenient feature of computing DIC by default, while in Stan no such features exist and the users have to write their own functions to compute DIC (e.g, Luo, 2019).[38] Estimation of IRT models with MCMC methods requires the specification of prior distributions for all model parameters, and we choose priors that are commonly seen in the Bayesian IRT literature. For the 3PLM model, we assign a standard normal distribution $N(0, 1)$ as the prior for the ability parameters for model identification, and a normal distribution with unknown mean and variance as the prior for the item difficulty parameter; the unknown mean is assigned a standard normal distribution $N(0, 1)$ as the hyperprior, and we assign the distribution $\gamma(1, 1)$ as the hyperprior for the precision parameter, which is the reciprocal of the variance. We assign a truncated normal distribution $N_{+}(0, 4)$ as the prior for item discrimination parameter, and a beta distribution $\beta(5, 23)$ for the pseudo-guessing parameter. For the testlet variance parameters in the 3PL TRT and 3PL CCTRT models, we assign as the prior normal distributions with a mean of zero and unknown precisions, for which $\gamma(1, 1)$ is assigned as the hyperprior.

4.4 Model Convergence Check

Since MCMC methods are used for model estimation, it is necessary to check whether model convergence has been reached before we draw inferences from the posterior distribution. In this study we apply the Gelman and Rubin's convergence diagnostic (Gelman & Rubin, 1992),[19] which computes the potential scale reduction factor (PSRF). PSRF values close to 1 indicate model convergence and as suggested by Gelman, Carlin, Stern, and Rubin (2014),[20] PSRF value of 1.1 can be used as the cutoff value to gauge model convergence in practice. For the 1PL and 2PL models, all PSRF values converge to 1 within 2,000 iterations, and we run three parallel chains with 4,000 iterations each to be conservative. For the 3PL, 3PL TRT and 3PL CCTRT models, we run three parallel chains with 5,000 iterations each and request every 10^{th} iteration to be used for inference (thinning = 10) to reduce autocorrelation in the posterior distribution for testlet variance parameters.

5. Results

5.1 Model Comparison

As can be seen from Table 2, the 3PL model has the smallest AIC, BIC, and DIC values among the three common dichotomous IRT models, indicating that it is the best fitting model. With the 3PL model as the base model, we estimated the corresponding 3PL TRT and the 3PL CCTRT models. Regarding the comparison among the three models, AIC, BIC, and DIC values consistently indicate that the 3PL CCTRT model has the best model fit, followed by the 3PL TRT model, and the 3PL model has the worst model fit. Using Anderson's suggestion (2008)[4] that a difference of nine or greater in those information criteria constitutes strong evidence for model choice, we find that the 3PL TRT model provides a model fit considerably better than the 3PL model with the differences in AIC, BIC, and DIC being 1226, 1187, and 670 respectively; the 3PL CCTRT model fits the data noticeably better than the 3PL TRT model with the differences in AIC, BIC, and DIC being 154, 149, and 50 respectively. In other words, when we model the item clustering effect due to passage effect in the 3PL TRT model, we find strong evidence that it should be chosen over the 3PL model, which assumes that no item clustering effect exists; when we model the dual item clustering effects due to passage effect and wording effect in the 3PL CCTRT model, model fit improves considerably over that of the 3PL TRT model, which assumes that no item clustering effect exists due to the NW format.

Table 2 Model Comparison Results

Model	AIC	BIC	DIC
1PL	86200	86431	87790
2PL	84877	85329	86430
3PL	84462	85135	85700
3PL TRT	83236	83948	85030
3PL CCTRT	83082	83799	84980

DOI: https://doi.org/10.30564/ jiep.v2i1.418

5.2 Magnitude of CIV

Table 3 lists the testlet variance estimates for the 3PL TRT and 3PL CCTRT models, which are indicative of the magnitude of CIV caused by the passage effects and negative wording effects. Xie (2014)[59] found in her simulation study that failure to model one source of item clustering effect in a cross-classified model results in biased testlet variance estimates, and with the increase of the magnitude of that item clustering effect, the bias increases. One natural question here is whether failure to model the wording effect leads to an incorrect interpretation of the magnitude of the item clustering effect due to passage effect. The comparison of the testlet variance estimates between these two models indicates that despite the better model fit that the 3PL CCTRT model has over the 3PL TRT model, such a model fit advantage does not translate into practical significances regarding the testlet variance estimates. The correlation between two sets of testlet variance estimates are greater than 0.99, and the values are nearly identical. As can be seen, none of the seven passages exhibits strong item clustering effect due to passage dependence, with the largest value being approximately 0.19. For the negative wording effect, the testlet variance estimate is 0.09 with a 95% credible interval not covering zero, indicating that only a small amount of CIV has been introduced into the test due to the NW format.

Table 3 Testlet Variance Estimates

Model	T_1 1	T_1 2	T_1 3	T_1 4	T_1 5	T_1 6	T_1 7	T_2 1
3PL TRT	0.13*	0.13*	0.15*	0.19*	0.11*	0.14*	0.18*	
3PL CCTRT	0.14*	0.13*	0.15*	0.18*	0.11*	0.14*	0.18*	0.09*

Note. * indicates that the 95% credible interval of the variance estimate does not cover zero.

5.3 Item and Ability Parameter Estimate Comparison

Table 4 Item Parameter Estimates Comparison

Item	Discrimination			Difficulty			Pseudo-Guessing		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
RC1	1.27	1.13	1.10	1.55	1.59	1.59	0.25	0.23	0.23
RC2	1.87	1.80	1.77	0.51	0.47	0.46	0.22	0.20	0.20
RC3	2.78	2.84	2.84	1.87	1.97	1.99	0.25	0.25	0.25
RC4	2.11	2.00	1.99	0.85	0.85	0.85	0.23	0.22	0.22
RC5	2.44	2.58	2.55	0.64	0.68	0.67	0.19	0.19	0.19
RC6	2.61	2.31	2.24	1.62	1.72	1.73	0.28	0.28	0.28
RC7	2.69	2.59	2.51	1.49	1.56	1.58	0.24	0.24	0.23
RC8	2.28	2.18	2.27	2.15	2.30	2.37	0.28	0.28	0.28
RC9	2.83	2.34	2.37	1.52	1.61	1.62	0.27	0.26	0.26
RC10	2.73	2.53	2.51	1.46	1.55	1.56	0.23	0.23	0.23
RC11	2.19	2.17	2.15	0.99	1.04	1.05	0.17	0.17	0.17
RC12	1.66	1.64	1.64	0.46	0.47	0.48	0.14	0.14	0.14
RC13	2.98	3.32	3.26	0.68	0.69	0.69	0.28	0.28	0.28
RC14	2.19	2.20	2.17	0.78	0.79	0.79	0.24	0.23	0.23
RC15	1.67	1.51	1.51	1.62	1.70	1.72	0.15	0.14	0.14
RC16	3.42	3.13	3.13	2.42	2.76	2.77	0.23	0.23	0.23
RC17	3.83	3.55	3.51	2.36	2.64	2.67	0.15	0.15	0.15
RC18	1.70	1.29	1.30	1.87	2.00	2.01	0.27	0.23	0.24
RC19	2.24	2.22	2.41	1.26	1.29	1.34	0.21	0.20	0.21
RC20	2.21	2.07	2.05	1.22	1.24	1.25	0.22	0.22	0.22
RC21	1.38	1.15	1.13	1.77	1.90	1.92	0.24	0.22	0.22
RC22	1.49	1.37	1.38	0.52	0.53	0.53	0.14	0.13	0.13
RC23	2.16	2.07	2.04	1.10	1.15	1.19	0.26	0.26	0.26
RC24	2.18	2.22	2.19	0.61	0.64	0.63	0.21	0.21	0.21
RC25	2.08	2.10	2.08	0.77	0.79	0.80	0.20	0.20	0.20
RC26	3.89	4.08	4.11	1.21	1.28	1.29	0.21	0.21	0.21
RC27	2.32	2.55	2.53	1.59	1.69	1.70	0.28	0.28	0.28
RC28	3.03	2.94	3.08	0.62	0.61	0.62	0.17	0.16	0.16
RC29	1.77	1.76	1.83	2.30	2.44	2.51	0.20	0.21	0.21
RC30	2.40	2.31	2.30	0.57	0.54	0.55	0.19	0.18	0.18
RC31	2.03	1.78	1.76	1.52	1.63	1.64	0.24	0.23	0.23
RC32	1.72	1.59	1.59	1.28	1.30	1.31	0.20	0.19	0.19
RC33	3.83	3.93	3.96	1.08	1.11	1.11	0.24	0.23	0.23
RC34	3.30	3.29	3.39	1.64	1.76	1.81	0.17	0.16	0.16
RC35	1.87	1.75	1.71	1.40	1.48	1.50	0.18	0.17	0.17
RC36	4.24	3.71	3.64	1.91	2.14	2.16	0.25	0.25	0.25
RC37	4.70	4.52	4.47	1.82	1.99	2.01	0.28	0.28	0.28
RC38	4.51	4.77	4.76	1.81	1.96	1.98	0.23	0.22	0.22
RC39	1.77	1.75	1.73	0.34	0.34	0.34	0.11	0.11	0.11
RC40	1.11	1.02	1.00	1.76	1.82	1.82	0.22	0.21	0.21

Note. M1 refers to the 3PL model, M2 the 3PL TRT model, and M3 the 3PL CCTRT model.

Table 4 lists the item parameter estimates from the three models. We observe that the three sets of item parameters are highly similar, although not identical. To further investigate whether there are any systematic patterns, we plot pairwise comparison of item and ability parameters and their corresponding standard errors in Figures 2-5. Note that the dotted line in these figures represents the regression line $y = x$, and how much a point deviates from this line indicates the magnitude of difference between two parameter estimates. Specifically, Figure 2 compares the differences in item parameter estimates and their standard errors between the 3PL and the 3PL TRT models. Such a comparison tells us the magnitude of differences regarding item parameter estimates as a result of not modeling the item clustering

 DOI: https://doi.org/10.30564/ jiep.v2i1.418

effect due to passage dependency. As can be seen, the two sets of item parameter estimates and their standard errors are similar with correlation values all greater than 0.96. Item discrimination parameter seems to be slightly overestimated in the 3PL model where the passage effect is not modeled, while its standard error appears to be slightly underestimated. Item difficulty parameter is virtually unaffected, while its standard error is somewhat underestimated. Item pseudo-guessing parameter is slightly overestimated, and its standard error is slightly underestimated.

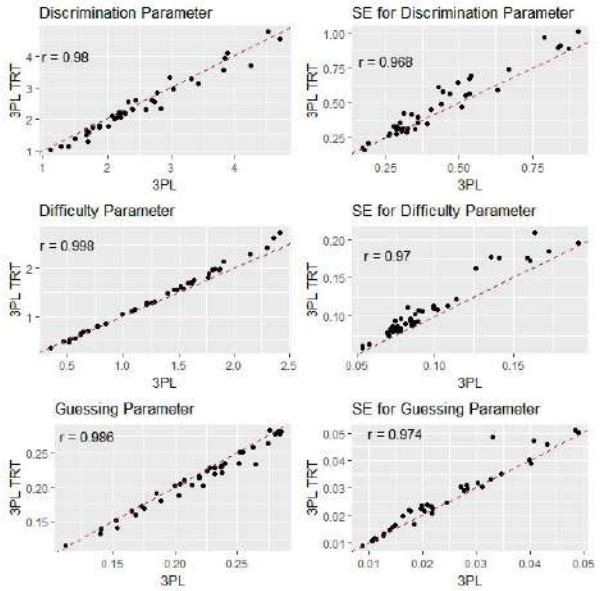
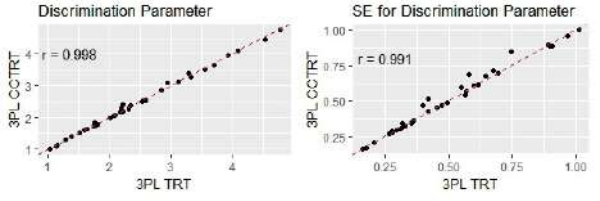


Figure 2. Item Parameter Comparison between the 3PL and 3PL TRT Models

Figure 3 compares the differences in item parameter estimates and their standard errors between the 3PL TRT and the 3PL CCTRT models. Such a comparison tells us the magnitude of differences regarding item parameter estimation as a result of modeling the item clustering effect due to passage dependency but not the item clustering effect due to the NW format. As can be seen, the two sets of item parameter estimates and their standard errors are nearly identical with the lowest correlation value being 0.991. In other words, item discrimination, difficulty, and pseudo-guessing parameters and their standard errors remain virtually the same when the item clustering effect due to the NW format is not modeled.



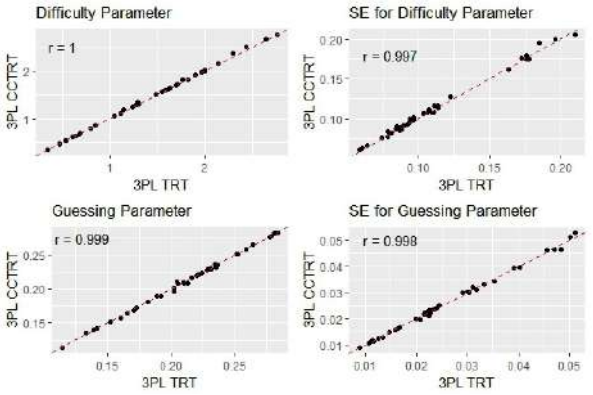


Figure 3. Item Parameter Comparison between the 3PL TRT and 3PL CCTRT Models

Figure 4 compares the differences in item parameter estimates and their standard errors between the 3PL and the 3PL CCTRT models. Such a comparison tells us the magnitude of differences regarding item parameter estimation as a result of not modeling the dual item clustering effects due to both passage dependency and the NW format. Similar to what is observed in Figure 2, item discrimination parameter is slightly overestimated and its standard error somewhat underestimated; item difficulty parameter is virtually unaffected and its standard error slightly underestimated; item pseudo-guessing parameter is slightly overestimated and its standard error virtually unaffected. However, it should be noted that the correlation values between two sets of parameter estimates and their standard errors are all extremely high with the lowest value being 0.968, which is the correlation value for standard errors of item discrimination parameter estimates.

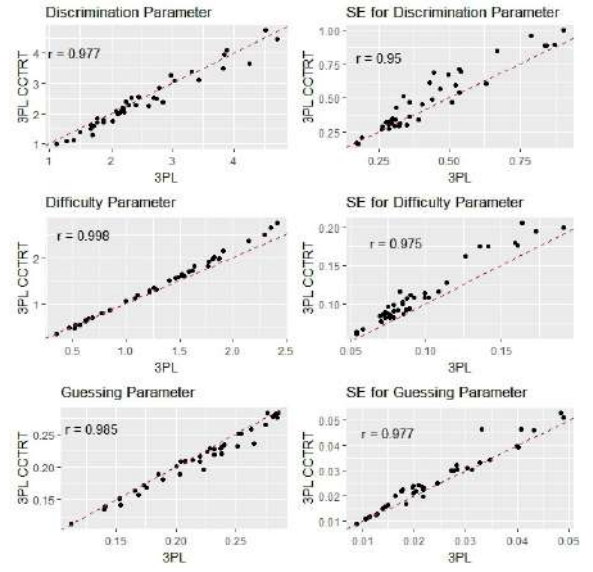


Figure 4. Item Parameter Comparison between the 3PL and 3PL CCTRT Models

DOI: https://doi.org/10.30564/ jiep.v2i1.418

As mentioned previously, the current data set was drawn from a high-stakes test that is used for admission and placement purposes. Consequently, it is critical to check whether the ability estimates and their standard errors are affected to an extent that would result in inaccurate estimation and erroneous classifications when different sources of CIV are not modeled or only partially modeled. Figure 5 provides such visual examinations. The top panel compares the differences in ability parameter estimates and their standard errors between the 3PL and the 3PL TRT models. As can be seen, the ability estimates are virtually unaffected with a correlation value of 0.998. The standard errors also seem to have a very high correlation value of 0.995, although it appears that standard errors of ability estimates from some examinees are slightly underestimated. The middle panel in Figure 5 compares the differences in estimated ability parameters and their standard errors between the 3PL TRT and the 3PL CCTRT models. Similar to what has been found in Figure 3 that item parameters are virtually the same between these two models, ability estimates and their standard errors are almost the same. The bottom panel in Figure 4 compares the differences in estimated ability parameters and their standard errors between the 3PL and the 3PL CCTRT models. Similar to the comparison between the 3PL and the 3PL TRT models, the ability estimates are virtually unaffected while some of their standard errors are slightly underestimated if item clustering effects due to the passage effect and the wording effect are not modeled.

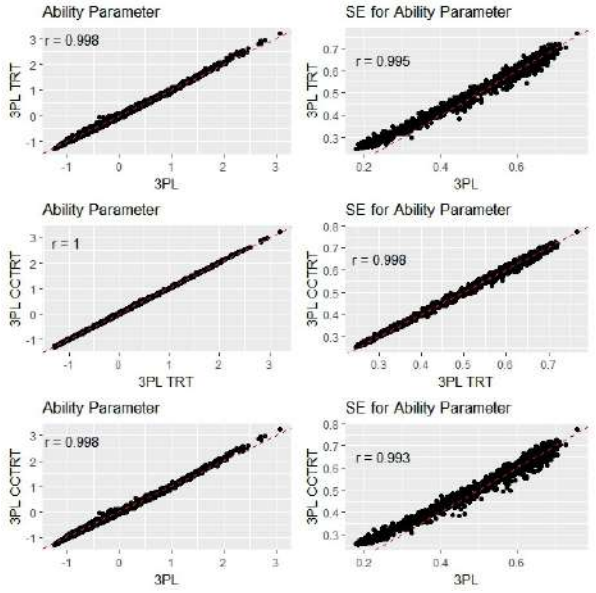


Figure 5. Ability Parameter Comparison between the Three Models

6. Discussion and Conclusions

The NW items have been extensively studied in the context of psychological instrument, especially through the lens of factor analysis to investigate whether they form a separate factor due to negative wording effect. In the educational testing literature, however, most studies focus on how the NW format affects item difficulty, and endeavors from the factor analysis perspective are scant, if not nonexistent. This study was intended to fill the gap in literature by investigating whether the NW items in a high-stakes English reading comprehension test form a separate factor and if yes, whether failure to model this particular factor leads to different item and ability parameter estimates.

Results indicate that similar to the findings in psychological instruments, the NW items in the current test does form a separate factor, of which the variance estimate is 0.09 with a 95% credible interval not covering zero. Note that the latent construct (English proficiency) that the current test is designed to measure is constrained to have a variance of one (for model identification purposes), which means that variance of the factor formed by NW items is less than one tenth of that of English proficiency. In simulation studies using the testlet model as a generating model, the variance of the testlet factor was usually generated to be 0.25 to represent small testlet effect (e.g., Li & Lissitz, 2012).[36] In this regard, it is reasonable to conclude that the NW factor has a small testlet effect.

We also investigated whether not modeling the NW factor leads to different parameter estimates. When the passage factor is modeled, not modeling the NW factor virtually makes no difference for the item and ability parameter estimates: the item and ability parameter estimates and their corresponding standard errors from the 3PL TRT and the 3PL CCTRT models have correlation values all greater than 0.99. If neither the passage factor nor the NW factor is modeled, the differences between the item and ability parameter estimates and their corresponding standard errors from the 3PL and the 3PL CCTRT models seem to be slightly bigger with all correlation values greater than 0.97, although we doubt that such differences would lead to any practical significance. The comparison of the parameter estimates between the 3PL and the 3PL TRT models also indicate that such differences between the 3PL and the 3PL CCTRT models are mainly due to the fact the passage factor is not modeled in the 3PL model.

The variance of the NW factor is small enough to be negligible. Consequently, the ability parameter and item parameter estimates are not noticeably affected. This finding is inconsistent with Wang, Chen, and Jin's finding

 DOI: https://doi.org/10.30564/ jiep.v2i1.418

(2015)[58] that for some of the NW items, the wording effect is large. Such a difference, we believe, should be attributed to the fact that the negative words used in the seven NW items in our data are both capitalized and bolded, while in their study they used a subscale of reading attitude assessment in Program for International Student Assessment (PISA) 2009 and two scales of math and science attitude assessment in Trends in International Mathematics and Science Study (TIMSS) 2011, in which the negative words within in NW items are not emphasized. Another possible cause is that among the seven NW items in our study, only two have negative words at the stem level and the other five at the option level (only one of the five items has the NW option as the correct answer). We believe that the negative words at the stem level has a more pronounced effect than at the option level, since misreading the negation at the stem level is more likely to result in an incorrect answer than at the option level. The results corroborate Haladyna's suggestion (2004)[22] that the negative words in a NW item need to be emphasized in that although the NW items with highlighted negative words introduce CIV into the current test, the magnitude of CIV is too small to cause differences of practical significance. We suspect that if the negative words in those NW items were not highlighted, the magnitude of the NW factor would be greater. However, to test such a hypothesis would require a real data set with such NW items, which are difficult, if not impossible, to find due to the popularity of Haladyna's advice.

We believe the CCTRT model can be a valuable validation tool in scenarios where more than one item clustering effect is expected to exist. Such scenarios may be common with educational testing data. For example, Baghaei and Aryadoust (2015)[5] analyzed a dataset drawn from responses to an English listening comprehension test consisting of 40 items with multiple item formats that fall under four listening passages. Suspecting that item format may constitute a source of CIV, they used a Rasch TRT model to account for such an item clustering effect and found that the testlet variance estimate for some format was large. Since their modeling approach does not account for the possibility that passage effect may form another source of CIV, the Rasch CCTRT model, we argue, is a more suitable model for their data that can simultaneous model both sources of CIV and hence produce more accurate parameter estimates. We recommend that in scenarios like this, a sensitivity analysis should be conducted to see whether the potential model improvement in the CCTRT model translates into differences in parameter estimates of practical significance. If not, we should proceed with the more parsimonious model despite its inferior model fit.

As shown previously, despite the fact that the 3PL TRT and 3PL CCTRT models have better model fit than the 3PL model, the three models lead to item and ability estimates that are highly similar and the most parsimonious one, the 3PL model, should be chosen as the ideal IRT model that combines model parsimony and practical utility. In this regard, this study can be regarded as a validation attempt to ascertain the underlying dimensionality for the data set used in the current analysis. Such an approach is similar to the bifactor approach to determining dimensionality advocated by Reise, Morizot, and Hays (2007),[46] who argue that the bifactor model always fits better than a one-factor model and hence, it is more informative to compare parameter estimates from the two models to see whether there is any difference of practical significance (if there is no noticeable difference between the two sets of parameters, then practical unidimensionality is established). Similarly, Luo and Al-Harbi (2016)[39] showed that when traditional dimensionality detection methods disagree, the bifactor approach can inform whether it makes any practical difference to proceed with the unidimensionality assumption.

To conclude, in this study we used a 3PL CCTRT model to investigate whether the NW items within in reading comprehension passages warrant modeling, and found that with the negative words highlighted, those NW items introduced CIV of negligible magnitude, a finding which supports the recommendation that if NW items have to be used, the negative words within should be accentuated. However, it should be emphasized that despite its small magnitude, the NW items do introduce into the current test CIV, which, regardless of its practical significance, still constitutes a threat to test validity. In addition, the current study used one single data set drawn from an English proficient reading comprehension test, and it is unknown whether the current findings are generalizable to the NW items used in other data sets or other tests that may or may not measure language proficiency. Consequently, we would like to recommend the judicious use of NW items in educational tests and reiterate the necessity for highlighting the negative words in cases where NW items have to be used.

Reference

[1] Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M. *Haladyna (Eds.), Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.

[2] Ackerman, T. A. (1987). *The robustness of LOGIST and BILOG IRT estimation programs to violations*

  DOI: https://doi.org/10.30564/ jiep.v2i1.418

of local independence. ACT Research Report Series, 87-14. Iowa City, IA: American College Testing.

[3] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory*, (pp. 267-281). Budapest, Hungary: Akademiai Kiado.

[4] Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence*. New York, NY: Springer.

[5] Baghaei, P., & Aryadoust, V. (2015). Modeling local item dependence due to common test format with a multidimensional Rasch model. *International Journal of Testing*, 15(1), 71-87.

[6] Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64(2), 153. doi:10.1007/bf02294533

[7] Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copula Functions for Residual Dependency. *Psychometrika*, 72(3), 393. doi:10.1007/S11336-007-9005-4

[8] Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. *American journal of pharmaceutical education*, 77(4), 71.

[9] Casler, L. (1983). *Emphasizing the negative: A note on the "not" in multiple-choice questions*. Paper presented at the meeting of the American Psychological Association.

[10] Cassels, J. R. T., & Johnstone, A. H. (1984). The effect of language on student performance on multiple-choice tests in chemistry. *Journal of Chemical Education*, 61, 613-615.

[11] Chen, W. H., & Thissen, D. (1997). Local Dependence Indexes for Item Pairs Using Item Response Theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265. doi:10.3102/10769986022003265

[12] Chessa, A. G., & Holleman, B. C. (2007). Answering attitudinal questions: Modelling the response process underlying contrastive questions. *Applied Cognitive Psychology*, 21,203-225. doi:10.1002/acp.1337

[13] Congdon, P. (2003). *Applied Bayesian modelling*. New York, NY: Wiley.

[14] Deemer, S. A., & Minke, K. M. (1999). An investigation of the factor structure of the teacher efficacy scale. *The Journal of Educational Research*, 93(1), 3-10.

[15] DeMars, C. E. (2006). Application of the Bi‐Factor multidimensional item response theory model to Testlet‐Based tests. *Journal of Educational Measurement*, 43(2), 145-168.

[16] Downing, S. M. (2005). The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education. *Advances in health sciences education*, 10(2), 133-143.

[17] Downing, S. M., Dawson-Saunders, B., Case, S. M., & Powell, R. D. (1991). *The psychometric effects of negative stems, unfocused questions, and heterogeneous options on NBME Part I and Part II item characteristics*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

[18] Dudycha, A. L., & Carpenter, J. B. (1973). Effects of item formats on item discrimination and difficulty. *Journal of Applied Psychology*, 58, 116-121.

[19] Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457-472.

[20] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis*. Boca Raton, FL, USA: Chapman & Hall/CRC.

[21] Gitchel, W. D., Roessler, R. T., & Turner, R. C. (2011). Gender effect according to item directionality on the perceived stress scale for adults with multiple sclerosis. *Rehabilitation Counseling Bulletin*, 55(1), 20-28.

[22] Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Erlbaum.

[23] Haladyna, T. M., & Downing, S. M. (1989a). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 37-50.

[24] Haladyna, T. M., & Downing, S. M. (1989b). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 51-78.

[25] Haladyna, T. M., & Downing, S. M. (2004). Construct‐irrelevant variance in high‐stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.

[26] Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, 15(3), 309-333.

[27] Harasym, P. H., Price, P. G., Brant, R., Violato, C., & Lorscheider, F. L. (1992). Evaluation of negation in stems of multiple-choice items. *Evaluation and the Health Professions*, 15, 198-220.

[28] Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological methods*, 2(3), 261.

 DOI: https://doi.org/10.30564/ jiep.v2i1.418

[29] Jiao, H., Wang, S., Wan, L., & Lu, R. (2009, April). *Investigation of local item dependence in scenario-based science assessment*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

[30] Jiao, H., Kamata, A., & Xie, C. (2015). A multilevel cross-classified testlet model for complex item and person clustering in item response modeling. In J. Harring, L. Stapleton, & S. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications*. Charlotte, NC: Information Age Publishing.

[31] Johnstone, A. H. (1983). Training teachers to be aware of the student learning difficulties. In P. Tamir, A. Hofstein A, & M. Ben Peretz (Eds.), *Preservice and Inservice Education of Science Teachers*. Rehovot (Isreal) – Philadelphia (USA): Balaban International Science Services.

[32] Kamata, A. (2001). Item Analysis by the Hierarchical Generalized Linear Model. *Journal of Educational Measurement*, 38(1), 79. doi:10.1111/j.1745-3984.2001.tb01117.x

[33] Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling*, 15(1), 136-153.

[34] Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style. *Quality & Quantity*, 47, 193-211. doi:10.1007/s11135-011-9511-4

[35] Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30(1), 3-21.

[36] Li, Y., & Lissitz, R. W. (2012). Exploring the full-information bifactor model in vertical scaling with construct shift. *Applied Psychological Measurement*, 36(1), 3-20.

[37] Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., & Raudsepp, L. (2012). Method effects:The problem with negatively versus positively keyed items. Journal of Personality Assessment, 94, 196-204. doi: 10.1080/00223891.2011.645936

[38] Luo, Y. (2019). LOO and WAIC as model selection methods for polytomous items. Psychological Test and Assessment Modeling, 61(2), 161-185.

[39] Luo, Y., & Al-Harbi, K. (2016). The Utility of the Bifactor Method for Unidimensionality Assessment When Other Methods Disagree. *SAGE Open*, 6(4).

[40] Luo, Y., & Jiao, H. (2018). Using the Stan program for Bayesian item response theory. Educational and psychological measurement, 78(3), 384-408.

[41] Luo, Y., & Liang, X. (2019). Simultaneously Modeling Differential Testlet Functioning and Differential Item Functioning: Addressing Variance Heterogeneity with a Multigroup One-Parameter Testlet Model. Measurement: Interdisciplinary Research and Perspectives, 17(2), 93-105.

[42] Magazine, S. L., Williams, L. J., & Williams, M. L. (1996). A confirmatory factor analysis examination of reverse coding effects in Meyer and Allen's affective and continuance commitment scales. *Educational and Psychological Measurement*, 56(2), 241-250.

[43] Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*. New York: American Council on Education and Macmillan

[44] Rachor, R. E., & Gray, G. T. (1996, April). *Must all stems be green? A study of two guidelines for writing multiple choice stems*. Paper presented at the annual meeting of the American Educational Research Association, New York.

[45] Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.

[46] Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19-31.

[47] Rijmen, F. (2010). Formal Relations and an Empirical Comparison among the Bi‐Factor, the Testlet, and a Second‐Order Multidimensional IRT Model. *Journal of Educational Measurement*, 47(3), 361-372.

[48] Rosenbaum, P. R. (1988). Items bundles. *Psychometrika*, 53(3), 349. doi:10.1007/bf02294217

[49] Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education, 35*, 117-134. doi:10.1080/02602930802618344

[50] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

[51] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.

[52] Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393-408.

[53] Tamir, P. (1991). Multiple choice items: How to gain the most out of them. *Biochemical Education*, 19(4), 188-192.

DOI: https://doi.org/10.30564/ jiep.v2i1.418

[54] Tamir, P. (1993). Positive and negative multiple choice items: How different are they? *Studies in Educational Evaluation*, 19, 311-325.

[55] Terranova, C. (1969). The effects of negative stems in multiple-choice test items. Dissertation Abstracts International, 30, 2390A.

[56] van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE, 8*(7), 1–7. http://doi.org/10.1371/journal.pone.0068967

[57] Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge: Cambridge University Press.

[58] Wang, W. C., Chen, H. F., & Jin, K. Y. (2015). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*, 75(1), 157-178.

[59] Xie, C. (2014). *Cross-classified modeling of dual local item dependence* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.

[60] Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of educational measurement*, 30(3), 187-213.

[61] Zhang, B. (2010). Assessing the accuracy and consistency of language proficiency classification under competing measurement models. *Language Testing*, 27(1), 119-140.

[62] Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLoS ONE*, 11(6), 1–15. http://doi.org/10.1371/journal.pone.0157795

 DOI: https://doi.org/10.30564/ jiep.v2i1.418

Author Guidelines

This document provides some guidelines to authors for submission in order to work towards a seamless submission process. While complete adherence to the following guidelines is not enforced, authors should note that following through with the guidelines will be helpful in expediting the copyediting and proofreading processes, and allow for improved readability during the review process.

Ⅰ. Format

- Program: Microsoft Word (preferred)
- Font: Times New Roman
- Size: 12
- Style: Normal
- Paragraph: Justified
- Required Documents

Ⅱ. Cover Letter

All articles should include a cover letter as a separate document.

The cover letter should include:

- Names and affiliation of author(s)

The corresponding author should be identified.

Eg. Department, University, Province/City/State, Postal Code, Country

- A brief description of the novelty and importance of the findings detailed in the paper

Declaration

v Conflict of Interest

Examples of conflicts of interest include (but are not limited to):

- Research grants
- Honoria
- Employment or consultation
- Project sponsors
- Author’s position on advisory boards or board of directors/management relationships
- Multiple affiliation
- Other financial relationships/support
- Informed Consent

This section confirms that written consent was obtained from all participants prior to the study.

- Ethical Approval

Eg. The paper received the ethical approval of XXX Ethics Committee.

- Trial Registration

Eg. Name of Trial Registry: Trial Registration Number

- Contributorship

The role(s) that each author undertook should be reflected in this section. This section affirms that each credited author has had a significant contribution to the article.

1. Main Manuscript

2. Reference List

3. Supplementary Data/Information

Supplementary figures, small tables, text etc.

As supplementary data/information is not copyedited/proofread, kindly ensure that the section is free from errors, and is presented clearly.

Ⅲ. Abstract

A general introduction to the research topic of the paper should be provided, along with a brief summary of its main results and implications. Kindly ensure the abstract is self-contained and remains readable to a wider audience. The abstract should also be kept to a maximum of 200 words.

Authors should also include 5-8 keywords after the abstract, separated by a semi-colon, avoiding the words already used in the title of the article.

Abstract and keywords should be reflected as font size 14.

Ⅳ. Title

The title should not exceed 50 words. Authors are encouraged to keep their titles succinct and relevant.

Titles should be reflected as font size 26, and in bold type.

Ⅳ. Section Headings

Section headings, sub-headings, and sub-subheadings should be differentiated by font size.

Section Headings: Font size 22, bold type

Sub-Headings: Font size 16, bold type

Sub-Subheadings: Font size 14, bold type

Main Manuscript Outline

Ⅴ. Introduction

The introduction should highlight the significance of the research conducted, in particular, in relation to current state of research in the field. A clear research objective should be conveyed within a single sentence.

Ⅵ. Methodology/Methods

In this section, the methods used to obtain the results in the paper should be clearly elucidated. This allows readers to be able to replicate the study in the future. Authors should ensure that any references made to other research or experiments should be clearly cited.

Ⅶ. Results

In this section, the results of experiments conducted should be detailed. The results should not be discussed at length in

this section. Alternatively, Results and Discussion can also be combined to a single section.

Ⅷ. Discussion

In this section, the results of the experiments conducted can be discussed in detail. Authors should discuss the direct and indirect implications of their findings, and also discuss if the results obtain reflect the current state of research in the field. Applications for the research should be discussed in this section. Suggestions for future research can also be discussed in this section.

Ⅸ. Conclusion

This section offers closure for the paper. An effective conclusion will need to sum up the principal findings of the papers, and its implications for further research.

Ⅹ. References

References should be included as a separate page from the main manuscript. For parts of the manuscript that have referenced a particular source, a superscript (ie. [x]) should be included next to the referenced text.

[x] refers to the allocated number of the source under the Reference List (eg. [1], [2], [3])

In the References section, the corresponding source should be referenced as:

[x] Author(s). Article Title [Publication Type]. Journal Name, Vol. No., Issue No.: Page numbers. (DOI number)

Ⅺ. Glossary of Publication Type

J = Journal/Magazine

M = Monograph/Book

C = (Article) Collection

D = Dissertation/Thesis

P = Patent

S = Standards

N = Newspapers

R = Reports

Kindly note that the order of appearance of the referenced source should follow its order of appearance in the main manuscript.

Graphs, Figures, Tables, and Equations

Graphs, figures and tables should be labelled closely below it and aligned to the center. Each data presentation type should be labelled as Graph, Figure, or Table, and its sequence should be in running order, separate from each other.

Equations should be aligned to the left, and numbered with in running order with its number in parenthesis (aligned right).

Ⅻ. Others

Conflicts of interest, acknowledgements, and publication ethics should also be declared in the final version of the manuscript. Instructions have been provided as its counterpart under Cover Letter.

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice

Aims and Scope

Journal of International Education and Practice is devoted to the study and exploration of education, reading, curriculum research, education policy, sociology of education, and economics.

It covers all innovations of educational research, including but not limited to:

- Basic education theories
- Education policies and regulations
- Education management
- Curriculum and teaching
- Examination and evaluation
- Education economy
- Education psychology
- Higher education
- International and comparative education
- Academic essays
- Education technology

Bilingual Publishing Co. (BPC)

Tel:+65 65881289

E-mail: contact@bilpublishing.com

Website: www.bilpublishing.com

About the Publisher

Bilingual Publishing Co. (BPC) is an international publisher of online, open access and scholarly peer-reviewed journals covering a wide range of academic disciplines including science, technology, medicine, engineering,education and social science. Reflecting the latest research from a broad sweep of subjects, our content is accessible worldwide – both in print and online.

BPC aims to provide an analytics as well as platform for information exchange and discussion that help organizations and professionals in advancing society for the betterment of mankind. BPC hopes to be indexed by well-known databases in order to expand its reach to the science community, and eventually grow to be a reputable publisher recognized by scholars and researchers around the world.

BPC adopts the Open Journal Systems, see on http://ojs.bilpublishing.com

Database Inclusion

Asia & Pacific Science Citation Index

Creative Commons

China National Knowledge Infrastructure

Google Scholar

Crossref

MyScienceWork

Bilingual Publishing Co. (UEN: 32307000X)
has maintained a prominent influence
in Southeast Asia for over 30 years.
Our mission is to advance academic research of the world
by publishing the latest research and encouraging discourse
across various scientific disciplines.
The main office of Bilingual Publishing Co
is located at 12 Eu Tong Sen Street, #07-169, Singapore 059819.

BPC
BILINGUAL
PUBLISHING CO.
Pioneer of Global Academics Since 1984

Tel: +65 65881289
E-mail: contact@bilpublishing.com
Website: www.bilpublishing.com

ISSN 2630-516X
01
9 772630 516190