JOURNAL OF BALTIC SCIENCE EDUCATION

ISSN 1648-3898

EVALUATING MEASUREMENT PROPERTIES OF ATTITUDINAL ITEMS RELATED TO LEARNING SCIENCE IN TAIWAN FROM TIMSS 2007

Pey-Yan Liou

Abstract. *In the recent empirical studies utilizing existing items and derived variables of international large-scale assessment (ILSA) data, the three major methodological deficiencies, including the use of a single item to define a construct, the statistical properties of ordinal data, and the fitness of the measurement structure for different scenarios, are examined. To overcome these issues, this study proposes an integrated approach to evaluating items and constructing derived variables in a given situation. Exploratory factor analysis, confirmatory factor analysis, and the item response model are utilized to evaluate student attitudinal items and derived variables from the Trends in International Mathematics and Science Study (TIMSS) 2007 Taiwanese data. The results suggest that the three-factor model composed of 12 items is optimal for the data, not the default factor structure in the database. The implications of evaluating items and creating derived variables from ILSA data for the education research community are also discussed.*

Key words: *attitudinal items, factor analysis, Trends in International Mathematics and Science Study.*

Pey-Yan Liou
National Central University, Taoyuan, Taiwan

Introduction

Due to the common interests of student science and mathematics education, international large-scale assessments (ILSA) have been launched and provide periodic data on student achievement as well as other related background information which allows the comparison of student science and mathematics achievement internationally. While mass communication media often report the ranking of a country based on the ILSA results, it is also essential for education researchers to extract specific diagnostic information which is necessary for international comparisons and improvements in education systems. Further analyses of ILSA data perhaps create the most influential knowledge base for education policy making in many countries (Olsen et al., 2011).

Students' attitudes towards science learning are one of the important issues, because they are viewed as a psychological belief which supports student learning processes, and are predictors of future career choice in the fields of science, technology, engineering and mathematics (Oliver & Simpson, 1988; Osborne et al., 2003). In ILSAs, items measuring students' attitudes are always surveyed due to their importance and relationship with achievement. Substantiating such associations between students' attitudes towards learning science and achievement is of paramount importance if education researchers wish to enlighten the policy debate. The Trends in International Mathematics and Science Study (TIMSS) as one of the major ILSAs serves as well-recognized data for researchers interested in examining relationships between students' attitudes towards learning science and achievement in both within- and between-country contexts.

The advantages for researchers conducting secondary data analyses using ILSA databases include the fact that they do not have to spend time and money collecting data, and they can gain access to large samples in either a single country or multiple countries. This, however, comes with the major disadvantage that the researchers do not have control over survey content. Researchers may have to utilize several existing related items in a questionnaire to construct a derived variable that they are interested in using in their studies (Bode, 1995). Researchers interested in the unobservable constructs, such as attitudes, may use several items to define the constructs which usually cannot be measured directly.

Students' attitudes towards learning science have been defined as individuals' beliefs about their academic characteristics and capabilities of learning science (Schunk, Pintrich, & Meece, 2008). Numerous researchers from different perspectives have proposed various definitions of students' attitudes. For instance, "self-concept" has been defined as people evaluating and judging their abilities and competences in school (Byrne & Shavelson, 1986; Marsh & Shavelson, 1985). On the other hand, based on the expectancy-value theories, "intrinsic motivation" indicates that people actively engage in an activity for the sake of their enjoyment and satisfaction in performing this activity. "Utility value" is defined as the usefulness of the task for the individual in terms of his/her future goals, and is related more to the ends of the task than to its means (Eccles et al., 1983; Wigfield et al., 1997). To study these attitudinal constructs and related issues, researchers usually combine several items from a questionnaire to create a derived variable. For example, many studies (e.g., Marsh, Kong, & Hau, 2001; Plucker & Stocking, 2001; Rinn, McQueen, Clark, & Rumsey, 2008) have averaged the values of the several items related to academic subjects in the Self-Descriptive Questionnaire II (Marsh, 1990), one of the most used questionnaires for measuring adolescents' attitudes, to form a derived academic self-concept variable. Taking Rinn, McQueen, Clark and Rumsey's study (2008) as an instance, ten items were utilized to form the derived variable, self-concept of learning mathematics.

While combining several items to form a derived variable is common practice in the current student attitude research, several published articles using previous TIMSS databases have only used a single item to indicate students' attitudes towards learning science. For instance, Shen and Tam (2008) used "I like science" as an indicator of self-perceived attitudes towards science, and "I usually do well in science" as self-efficacy. In Wilkins' article (2004), he employed "I usually do well in science" as an expression of self-concept. While using only a single item is straightforward, combining information into a composite may be preferable. As Wilkins (2004) cautioned, "The inability to estimate the reliability of the self-concept measures, because they were created from a single item, did not allow for the elimination of the possibility that differences across countries resulted from measurement error" (p. 345). While studies focused on students' attitudes towards learning often use a single item to define such a broad construct, researchers wishing to use TIMSS data should utilize the combining strategy to create a derived variable to improve measurement precision, scope and validity.

In fact, several derived variables related to students' attitudes towards learning science were created in the TIMSS 2007 database, such as "the index of students' positive affect toward science." The sum and average method was used to create composite scales from the original items. This derived variable was created by summing the values of the three items and using the average as the value for the derived variable. The three items are, "I enjoy learning science," "science is boring," and "I like science." The response categories for the three rating-scale items are "agree a lot (=1)," "agree a little (=2)," "disagree a little (=3)," and "disagree a lot (=4)." The negatively worded items were reverse coded. The value of this derived variable was computed across the three items; that is, a high level indicates an average score of less than or equal to 2, a medium level indicates average scores of greater than 2 but less than 3, and a low level indicates an average score equal to or greater than 3 (Martin & Preuschoff, 2008).

Using the derived three-point rating scale attitudinal variables in the TIMSS 2007 database seems to overcome the issue of measurement reliability caused by using a single item. However, the Likert-type variables do not meet the assumption of a parametric statistical test, that is, that the data are equal-interval, normally distributed, and of equal variance (Liu & Boone, 2006). While using non-parametric statistical techniques may be a solution, Harwell and Gatti (2001) also stated that rescaling the Likert-type data into interval data is the most attractive and practical method for solving the dilemma of coupling measurement scales with statistical analyses. Thus, researchers should seriously consider creating their own derived variables with a continuous rather than an ordinal property with limited points when using rating scale derived variables from ILSA data.

Additionally, the three derived attitudinal variables, formed from 11 items via factor analysis, in the TIMSS 2007 database were based on all students' responses from all participating education systems. The factor structure of these derived variables may not, however, be optimal for data which are only from one country (Eklöf, 2007; Liu & Meng, 2010; Sabah, Hammouri, & Akour, 2013). If researchers are interested in using data from only one country or from certain countries, the three-factor structure composed of 11 items measuring students' science attitudes may not fit the data well. Therefore, education researchers may be interested in using different methods to create derived attitudinal variables with solid statistical evidence to meet their needs. While ILSAs provide numerous data for public use, researchers should be aware of several statistical methodology issues related to the released data

ISSN 1648–3898

and how to make better use of them. Thus, this paper aims to propose the use of a series of approaches to evaluate items and to construct derived variables regarding students' attitudes towards learning science in Taiwan.

Research Purposes

Researchers (Eklöf, 2007; Liu & Meng, 2010; Marsh et al., 2013) have suggested that more studies need to be conducted to show the validity of the factor structure of the attitudinal items in a given situation (e.g., when data from a single country are used) while utilizing TIMSS data. Additionally, the evaluation results may serve as feedback for researchers interested in using TIMSS attitudinal items for their future studies. Therefore, the following two research questions were addressed using the Taiwanese portion of the TIMSS 2007 data:

1. How does the three-factor model formed from the 11 items related to students' attitudes towards learning science provided in the TIMSS 2007 database work for eighth-grade Taiwanese students?
2. What are the psychometrical properties of the items and emerging attitudinal derived variables from the TIMSS 2007 eighth-grade Taiwanese student data?

Methodology of Research

The methodology section contains three subsections: a sample and data source section, a section containing main measures used in this study, and a section on the statistical techniques employed. In the sample and data source section, an overview of the data from the Taiwanese portion of TIMSS 2007 and the sampling scheme is given. In the measures section, the 12 items regarding students' attitudes towards learning science are examined. Finally, a series of analytical methods, including factor analyses and item response modeling are presented and used to answer the research questions.

Sample and Data Source

Eighth-grade students in Taiwan from TIMSS 2007 were selected as the sample in this study. It is well known that the students in East Asian countries, including Taiwan, Hong Kong, Japan, and South Korea, performed very well in these ILSAs; however, their attitudes towards learning are negative compared with those of students from other countries (Martin, Mullis, & Foy, 2008). Many researchers (e.g., Liu & Meng, 2010; Shen & Tam, 2008) have stated a need to shed further light on these students' attitudes towards learning and further to compare the differences with Western countries where most of the theories of student attitudes and achievement originated.

In order to sample represented students in each country, TIMSS 2007 used a two-stage stratified cluster sampling design (Olson, Martin, & Mullis, 2008). Due to the stratified cluster sampling, weighting must be applied to ensure accurate representation of the population when analyzing large-scale databases (Liou & Hung, in press). Total student weight was used to take student sampling weight into consideration in this study. Thus, there were 4,046 such students in the database from 150 schools, and the sum of the weighted students was 307,288.

Measures

Twelve items were used to construct the derived variables related to students' attitudes towards learning science. They are 1) I usually do well in science, 2) I would like to take more science in school, 3) Science is more difficult for me than for many of my classmates, 4) I enjoy learning science, 5) Science is not one of my strengths, 6) I learn things quickly in science, 7) Science is boring, 8) I like science, 9) I think learning science will help me in my daily life, 10) I need science to learn other school subjects, 11) I need to do well in science to get into the <university> of my choice, and 12) I need to do well in science to get the job I want. The response categories for these items were "agree a lot (=1)," "agree a little (=2)," "disagree a little (=3)," and "disagree a lot (=4)." Although only 11 of these 12 items were used to create the three derived variables in the TIMSS 2007 database, all of the 12 items were utilized in this study. The 11 items forming the three derived variables may work best for the data from all participating countries; however, this may not be the best solution for data from only one country. Responses to positive statements were reverse coded as a higher score means more positive perceptions of attitudes towards learning science.

Additionally, five plausible science achievement scores were used to correlate with and validate the derived attitudinal variables. In order to measure a broad coverage of science curricula topics, a complex matrix-sampling booklet design is used in TIMSS assessment. Plausible values were randomly drawn from the distribution of ability estimates that represent the range of reasonable values for a student's ability (Foy, Galia, & Li, 2008).

Data Analyses

Three approaches were used in this study, namely exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and the multidimensional random coefficients multinomial logit model (MRCMLM). Finally, Cronbach's alpha was calculated to show the reliability of each scale, and the correlations between each scale and student science achievement are presented to show the validation of the scale.

EFA was utilized to identify underlying factors among the items. CFA was then performed to make model comparisons to confirm the factor structure and to provide additional data regarding the patterns that emerged. The software programs SPSS (2008) and MPlus (Muthén & Muthén, 1998-2010) were used for EFA, and MPlus was used for CFA. As it is not appropriate to specify a CFA model based on the results of an EFA and to obtain the estimates using the same data, the 4,046 samples were randomly divided into two datasets, of which 2,023 samples were used for EFA, and 2,023 were used for CFA to confirm the results.

After determining the factor structure from CFA, the MRCMLM (Adams, Wilson, & Wang, 1997), an extension of the Rasch family of IRT models, was applied to analyze the data. MRCMLM provides estimates of the correlations among the derived variables and yields unbiased estimates of item parameters. The use of the MRCMLM was not only to transform the separate rating scale items into a continuous derived variable, but also to provide a fuller description of the items in each derived variable (Bode, 1995; Reeve & Fayers, 2005). This occurs as the MRCMLM transforms the ordinal raw scores from the items into equal-interval logit units which can represent a linear relationship between items and persons on the same scale. The software program ConQuest (Wu, Adams, & Wilson, 2007) was used to perform the analysis. The Rating Scale Model (RSM; Andrich, 1978; Wright & Masters, 1982) was adopted.

Furthermore, Cronbach's alpha was calculated in SPSS (2008) for each set of items defining a separate scale. Further, after the final structure for the items was decided, factor scores for the derived attitudinal variables were saved and regressed on the five plausible science scores. IDB Analyzer (2009) was used to analyze the relationship between derived attitudinal variables and science achievement. IDB Analyzer accommodates the five plausible values of the science achievement scores into the analyses, and results are aggregated to yield accurate estimates and standard errors which incorporate sampling and imputation errors.

Results of the Research

Exploratory Factor Analysis (EFA)

The correlation matrix for the twelve items for the 154,225 weighted samples is shown in Table 1. The Kaiser-Meyer-Olkin measure of sampling adequacy index was .91 and Bartlett's test of sphericity was significant (p <0.001), indicating that the sample and correlation matrix were appropriate for the analysis. Based on the eigenvalue-greater-than-one rule, a two-factor structure should be sufficient to represent the 12 items (see Table 2). Table 3 shows the pattern of factor loading and communality of each item. There is a clear pattern that eight items are located in Factor 1, while the other four are in Factor 2. The communality of each item is larger than .5. The first two factors explain about 64% of the variance in the 12 items. Note, however, that the eigenvalue-greater-than-1 rule has been criticized by Zwick and Velicer (1982) as this rule of thumb is often considered as overestimating the factors. Additionally, the result shows the large dominance of the first dimension. Given the discrepancy between the eigenvalues for the first and second dimensions, one can err on the side of parsimony and decide that there is only one factor, since the second dimension accounts for a relatively small amount of variance compared to the first. Therefore, both the two-factor and one-factor models are considered in the subsequent CFA analysis.

Table 1. Intercorrelations, means, and standard deviations of the 12 items from the EFA and CFA datasets.

Item	1	2	3	4	5	6	7	8	9	10	11	12	*M*	*SD*
1		.53	.40	.63	.52	.67	.45	.63	.40	.34	.40	.39	2.50	.85
2	.54		.26	.65	.35	.52	.51	.64	.45	.44	.45	.43	2.39	.93
3	.46	.27		.37	.60	.42	.44	.37	.17	.16	.17	.19	2.36	.94
4	.62	.66	.39		.49	.66	.62	.82	.49	.41	.47	.45	2.46	.93
5	.54	.34	.57	.46		.49	.51	.48	.24	.22	.24	.25	2.23	1.00
6	.67	.51	.45	.66	.50		.47	.67	.41	.38	.43	.40	2.26	.85
7	.45	.51	.44	.62	.49	.47		.66	.40	.31	.34	.33	2.68	.97
8	.62	.66	.40	.84	.44	.67	.66		.49	.42	.47	.45	2.47	.96
9	.42	.45	.20	.49	.25	.41	.40	.49		.53	.45	.46	3.00	.84
10	.37	.44	.15	.45	.22	.38	.31	.42	.53		.54	.53	2.36	.84
11	.39	.43	.19	.46	.24	.43	.34	.47	.45	.54		.74	2.58	.97
12	.37	.44	.17	.45	.24	.40	.33	.45	.46	.53	.74		2.38	.95
M	2.50	2.41	2.37	2.49	2.23	2.25	2.66	2.48	3.04	2.36	2.59	2.40		
SD	.84	.93	.94	.94	1.02	.85	.97	.96	.84	.84	.99	.95		

Note. 1= I usually do well in science; 2= I would like to take more science in school; 3=Science is more difficult for me than for many of my classmates; 4=I enjoy learning science; 5=Science is not one of my strengths; 6=I learn things quickly in science; 7=Science is boring; 8=I like science; 9=I think learning science will help me in my daily life; 10=I need science to learn other school subjects; 11=I need to do well in science to get into the <university> of my choice; 12=I need to do well in science to get the job I want; M= Mean; SD =Standard Deviation; the lower diagonal shows the values for the EFA data with N=154,225 and the upper diagonal is for the CFA data with N=153,063.

Table 2. The initial eigenvalues, percentages of variance, and cumulative percentages for factors of the 12 items regarding student attitudes toward learning science.

Factor	Eigenvalue	% of Variance	Cumulative %
1	6.10	50.81	50.81
2	1.56	13.00	63.81
3	0.83	6.89	70.70
4	0.65	5.38	76.09
5	0.59	4.88	80.96

Table 3. Principal axis factor analysis of student attitudinal items related to learning science items with oblique promax rotation.

Items	F1	F2	h2
Factor 1: Students' Self-concept and Intrinsic Motivation of Learning Science			
I usually do well in science	**0.70**	0.10	0.59
I would like to take more science in school	**0.41**	0.38	0.51
Science is more difficult for me than for many of my classmates	**0.79**	-0.28	0.41
I enjoy learning science	**0.65**	0.27	0.73
Science is not one of my strengths	**0.80**	-0.19	0.47
I learn things quickly in science	**0.70**	0.11	0.61
Science is boring	**0.68**	0.04	0.51

Items	F1	F2	h2
I like science	**0.66**	0.27	0.74
Factor 2: Students Utility-Value of Learning Science			
I think learning science will help me in my daily life	0.10	**0.60**	0.45
I need science to learn other school subjects	-0.08	**0.77**	0.53
I need to do well in science to get into the <university> of my choice	-0.11	**0.84**	0.60
I need to do well in science to get the job I want	-0.16	**0.89**	0.63
Eigenvalue	5.69	1.09	
Cumulative percent of variance explained	47.42	56.50	

Note. h^2=communality

Confirmatory Factor Analysis (CFA)

The correlation matrix for the twelve items of 12,226 weighted samples is shown in the upper diagonal of Table 1. CFA was conducted to make model comparisons and to examine item patterns among factors. The one-factor, two-factor, three-factor composed of 11 items (the default in TIMSS 2007), and the three-factor composed of 12 items models were examined with CFA. Furthermore, an EFA with a fixed number of factors set as 3 to be extracted was conducted to examine the patterns of the 12 items due to the default model composed of three derived variables in TIMSS 2007. The result showed that four items are located in each of the three derived variables, so this model was also adopted in this study. By comparing the four models, there may be more statistical evidence to show which model better fits the data.

Table 4 summarizes the degrees of freedom, Chi-square, AIC, BIC, CFI, and TLI for the four models. After taking the degrees of freedom into consideration, the Chi-square deviance test indicates that the three-factor model composed of 12 items has a better fit. The AIC and BIC of this model also have the lowest values compared with the other models. Additionally, the results of the other three fit indices (i.e., CFI, TLI, and RMSEA) support the three-factor model composed of 12 items as being superior to the other models. The values of CFI and TLI of the three-factor model with 12 items are higher than those of the other models, while the value of RMSEA is lower than that of the other models. Moreover, according to the model-comparison approach proposed by Chen (2007), the more constrained model (i.e., the three-factor model composed of 12 items) is considered as the better model if a change in CFI for this model is greater than .01 and RMSEA shows as a better fit than those for the less constrained model. In sum, based on these results, the three-factor model composed of 12 items was chosen as the best model, and the two-factor model is the second best. The three-factor model composed of 11 items, the default in TIMSS 2007, has a worse fit than the other two models. The one-factor model is the worst of the four models. Therefore, the three-factor model composed of 12 items was selected for further examination in terms of CFA structure, item fit, item map, and reliability analysis.

Table 4. Goodness-of-fit indices for four models of the CFA dataset.

Model	*df*	Chi Square	AIC	BIC	CFI	TLI	RMSEA
One Factor	54	2626.66	52533.26	52734.44	0.81	0.77	0.16
Two Factor	53	1466.17	51374.77	51581.54	0.89	0.87	0.12
Three Factor (11 items)	52	2308.54	52219.15	52431.50	0.83	0.79	0.15
Three Factor (12 items)	51	1047.03	50959.63	51177.58	0.93	0.90	0.10

The details of the three-factor model composed of the 12 items (standardized regression weights, the squared multiple correlations) are reported in Table 5.When Factor 3 increases by one standard deviation, the item, "I think learning science will help me in my daily life," increases by .61 standard deviations after controlling for the other three items, "I need science to learn other school subjects,""I need to do well in science to get into the <university> of my choice," and "I need to do well in science to get the job I want." Meanwhile, the squared multiple correlation coefficients can be interpreted as the proportion of the item variance that is explained by the common factor. The

ISSN 1648–3898

remaining percentage of its variance is accounted for by the unique factor (error). The squared multiple correlation coefficients can be interpreted as follows: considering the item, "I usually do well in science," as an example, 64% of its variance is accounted for by Factor 1. The remaining 36% of its variance is accounted for by the unique factor (error).

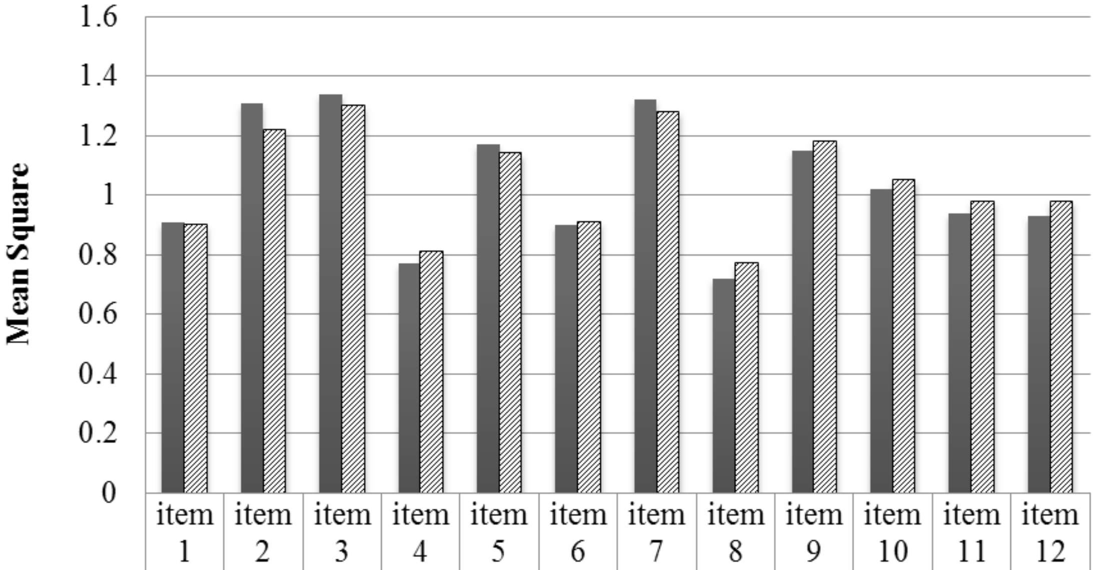
As for the relationship between the three factors, the results indicate that there is a positively moderate to high relationship (.60~.86) between them. The correlation coefficient between the first factor and the second factor is high (.86), between the first and the third it is moderate (.60), and between the second and third it is moderate (.68). According to the meaning of these items in each factor, the first factor is called "Students' Science Learning Self-Concept," based on the definition of self-concept provided by Marsh and Shavelson (1985), the second is called "Students' Intrinsic Motivation in Learning Science," based on Wigfield et al. (1997) who referred to the enjoyment aspect of task interest, and the third factor is called "Students' Utility-Value of Learning Science," based on the task value beliefs of expectancy-value theory (Eccles et al., 1983).

Table 5. Standardized regression weights, squared multiple correlations and Cronbach's alphas for the three-factor model composed of 12 items in CFA.

Item	Standardized regression weights	Squared multiple correlations
Factor 1: Students' Science Learning Self-concept (alpha= 0.81)		
I usually do well in science	0.80	0.64
Science is more difficult for me than for many of my classmates	0.55	0.30
Science is not one of my strengths	0.66	0.43
I learn things quickly in science	0.83	0.68
Factor 2: Students' Intrinsic Motivation in Learning Science (alpha= 0.88)		
I would like to take more science in school	0.72	0.51
I enjoy learning science	0.89	0.80
Science is boring	0.71	0.50
I like science	0.91	0.82
Factor 3: Students Utility-Value of Learning Science (alpha= 0.83)		
I think learning science will help me in my daily life	0.61	0.37
I need science to learn other school subjects	0.69	0.47
I need to do well in science to get into the <university> of my choice	0.83	0.69
I need to do well in science to get the job I want	0.80	0.64

Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM)

The MRCMLM section examines the evidence for construct validity in terms of item fit and the person-item fit indicated by an item map. For evaluating item fit, the unweighted fit mean square statistics (UFMS) and the weighted fit mean square statistics (WFMS) were used to indicate the item fit. For survey rating scale items, an outfit value of between 0.60 and 1.40 is considered as an acceptable range (Wright & Linacre, 1994). An outfit value less than 0.60 suggests that an item does not contribute information to the test beyond that provided by the rest of the items. A value larger than 1.40 indicates that an item does not define the same construct as do the rest of the items. Based on the results of this study, the mean square fit values for all items are all located in the range of 0.60 - 1.40, so these items were considered to provide enough information to their factor (see Figure 1).

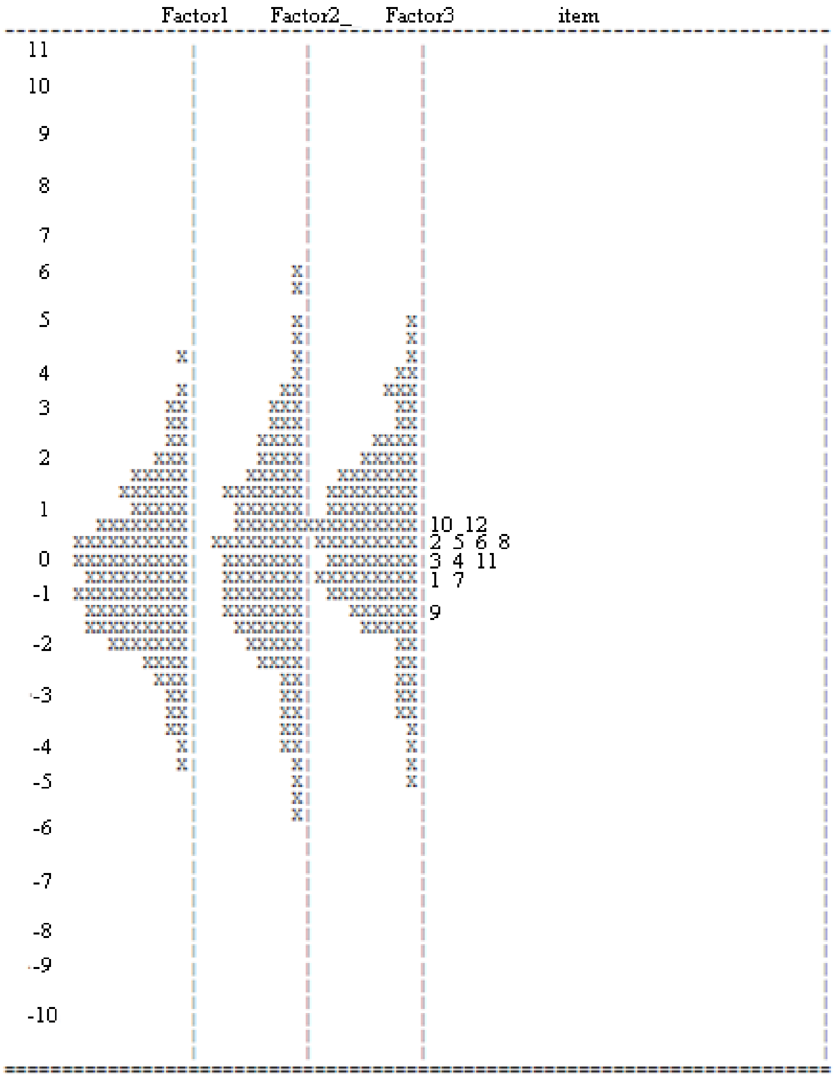


	item 1	item 2	item 3	item 4	item 5	item 6	item 7	item 8	item 9	item 10	item 11	item 12
UFMS	0.91	1.31	1.34	0.77	1.17	0.9	1.32	0.72	1.15	1.02	0.94	0.93
WFMS	0.9	1.22	1.3	0.81	1.14	0.91	1.28	0.77	1.18	1.05	0.98	0.98

Figure 1: Illustration of the UFMS and WFMS measures per item.

Figure 2 shows the item map for the three-factor model with the 12 items. The item map includes the student belief distribution and the item location distribution, and represents person and item estimates on the same metric to allow for direct comparisons. For the Rasch-type model, the probability of rating an item is viewed as a function of the student belief level and the item location level. The larger the logit value, the higher the student estimate's belief level. An item with a higher value is less likely to be endorsed. The further a student's belief estimate is above the item location estimate, the more likely it is that the student rates the item more highly. The item distribution should cover the span of the student belief estimates, so that the item can present accurate measures for students of all belief levels. In this case, the student distributions spanned around 9 logits with the logit of 0 for the median in Factor 1, around 12 logits for Factor 2, and around 10 logits for Factor 3. The range of student beliefs is wide, but not as wide as for the three factors composed of four items. Thus, if students have extremely high latent attitudes, these items may not measure the constructs well.

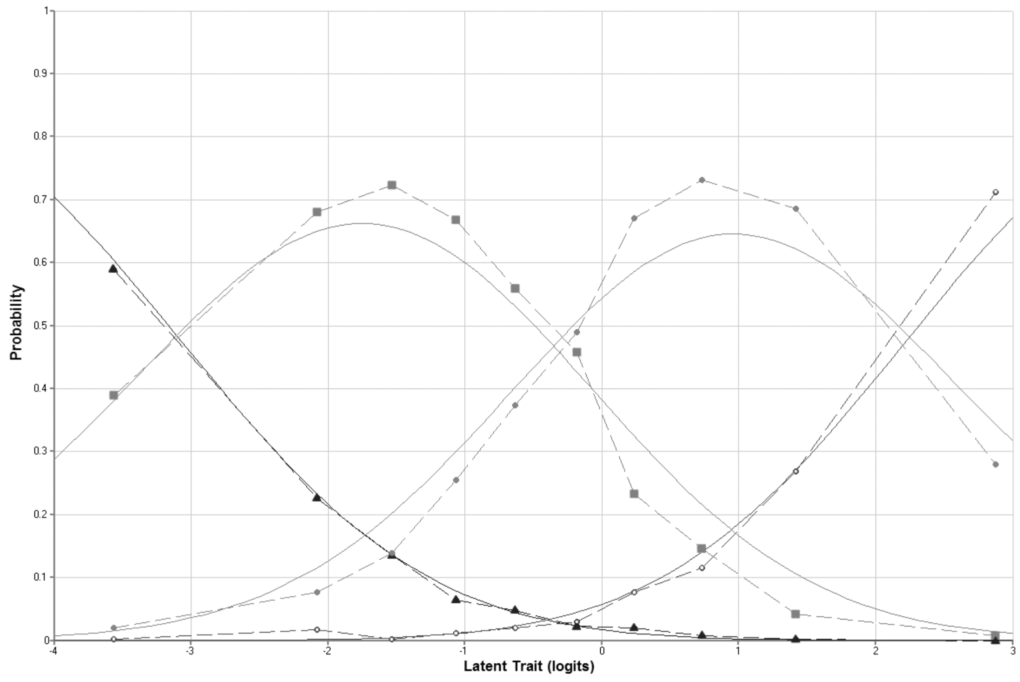
ISSN 1648–3898



Note. For each scale, X stands for 35.6 cases. 1= I usually do well in science; 2= I would like to take more science in school; 3= Science is more difficult for me than for many of my classmates; 4= I enjoy learning science; 5= Science is not one of my strengths; 6= I learn things quickly in science; 7= Science is boring; 8= I like science. 9= I think learning science will help me in my daily life; 10= I need science to learn other school subjects; 11= I need to do well in science to get into the <university> of my choice; 12= I need to do well in science to get the job I want.

Figure 2: Wright map for the three-factor model with 12 items.

Additionally, the function of the response format was visually assessed by examining the category response function (CRF) for each item. The CRF graphs for items 1 and 9 are shown in Figures 3-4. The CRF shows the probability of selecting the category at each logit along the Rasch scale continuum. The range on the Rasch scale continuum where a curve is higher than other curves in the graph signifies where a response category has a greater probability of being selected than all other response categories. Taking Figure 3 of item 1 for instance, students with a 1 belief level were likely to agree, as their probability rating "agree a little" was around .65. Similarly, students with a 1 belief level were likely to disagree, as their probability rating "disgree a little" was around .60. Moreover, according to Figure 3 of item 1, the four responses for the item were distinguished quite evenly, meaning that the four response formats functioned well. However, the response format function for item 9 may not perform as well as it does for item 1 because options one and two can only measure students with a low level of belief (-1.5 logits).



Note. The smooth line indicates the modeled item characteristic curve and the broken line shows the empirical item characteristic curve.

Figure 3: CRF for item 1.

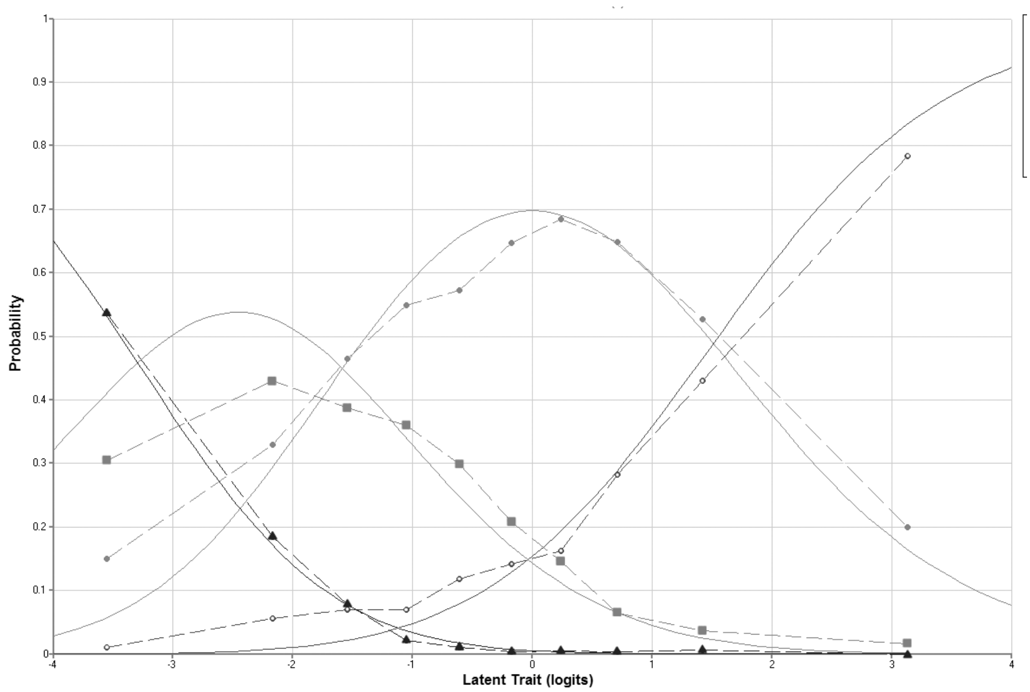


Figure 4: CRF for item 9.

ISSN 1648–3898

Reliability Analysis and Derived Variable Validation

The Cronbach's alpha for each scale is also presented in Table 5. The Cronbach's alpha of the three derived variables is greater than 0.8, so they can be considered to meet the criteria for group level decisions, thereby supporting our decision to retain these items and factors for future research. As for the derived variable validation, the correlations between the student attitudinal factor scores and science achievement are all positive. The percentage of variance in student science achievement accounted for by Factor 1 is .14, by Factor 2 it is also .14, and by Factor 3 it is .10.

Discussion

The results of this investigation show that the default TIMSS three factor structure composed of the 11 items does not work well for the TIMSS 2007 Taiwanese eighth grade student data. The two-factor model from the 12 attitudinal items initially emerged from EFA. A series of CFA based on different models was conducted for model comparison. Based on several fit indices, the results of the CFA indicated that the three-factor model composed of the 12 items works better than the other models, and even the two-factor model works better than the default TIMSS model. While the evidence from the Cronbach's alphas and the moderate correlations between the derived variables and science achievement is sufficient, the results of IRT reveal that too few items are available to provide an accurate measure for students whose attitudes are either high or low.

In surveying the current practice of studies in which students' attitudes towards learning as captured in TIMSS data were utilized, most of the studies did not utilize any statistical techniques to fully examine the items and derived variables. These studies either used a single item (e.g., Wilkins, 2004) or directly utilized a default Likert-type derived variable (e.g., Kaya & Rice, 2010) to represent students' attitudes. Liu and Meng (2010), Eklöf (2007), and Sabah, Hammouri, and Akour (2013) are among the very few studies in which the issue of the factor structure of the attitudinal items in TIMSS data is paid attention to. In Liu and Meng's study (2010), TIMSS 2003 eighth grade student data from Japan, Hong Kong, Taiwan and the U.S. were selected for analysis. EFA was utilized to examine 12 student attitudinal items. Their results showed that a two-factor structure composed of six items each emerged from the data, while the default structure of TIMSS 2003 data is the other two-factor structure, one with five items and the other with seven items. Eklöf (2007) first utilized EFA to examine the 12 attitudinal items from TIMSS 2003 Swedish eighth grade student data. The result of EFA indicates that a two-factor model was shown. Further, CFA was conducted to compare two models (one is the two-factor model derived from EFA, and the other is based on theoretical assumptions), and showed that the four-factor model fitted the data slightly better than the two-factor model. Both of these two studies highlight the need to examine the factor structure in a given situation, and Eklöf (2007) further utilized CFA to compare models.

The current investigation and both Liu and Meng's (2010) and Eklöf's (2007) studies indicate that using the default derived variables and factor structure in TIMSS data needs further consideration. Both EFA and CFA serve as practical and complementary tools to evaluate the properties of the items and derived variables. While EFA is an exploratory approach to determining the number of derived variables, CFA is another means to compare and validate models. For both this study and Eklöf's study (2007) in which EFA and CFA were utilized, the results show that there is a discrepancy between the EFA and CFA results. While EFA is a data-driven technique to uncover the latent dimensions of items, CFA can further compare and confirm the models which emerge from either EFA or prior theories as setting the pattern of items and derived variables fixed. Both the results of EFA and CFA provide evidence for researchers to make appropriate decisions regarding how to evaluate items and to form valid derived variables.

In this study, based on the contents of the items shown in each factor, three factors are named as "Students' Science Learning Self-Concept," "Students' Intrinsic Motivation in Learning Science," and "Students' Utility-Value of Learning Science," respectively. This result corroborates the many studies (e.g., Martin, Mullis, & Foy, 2008; Valentine, DuBois, & Cooper, 2004) which have shown that the relationship between students' attitudes towards learning and achievement are positively correlated. Valentine, DuBois, and Cooper (2004) conducted a meta-analysis to examine the relation between self-beliefs and academic achievement based on 55 studies including evaluations of 282 separate effect sizes. The average effect size of these surveyed studies is .09. The range of the effect sizes of these studies is from -.01 to .36, and most are positive. In the current investigation, the correlations between the three attitudinal derived variables and achievement are all around .35, which are relatively high compared with other studies.

Meanwhile, it is important to notice that Taiwanese students' science achievement has higher positive relationships with their self-concept and intrinsic motivation than utility-value. While students' attitudes towards learning science in general have significant positive relationships with science achievement, the results show that students' self-concept and intrinsic motivation explain more variance of student achievement than utility-value. Therefore, science education researchers and practitioners should consider designing and implementing curricula to bolster students' self-concept of and intrinsic motivation to learn science in particular. For instance, developing meaning, relevance, and an integrated-system approach to the science curriculum in middle schools is suggested (e.g., Chang, 2005; Singh et al., 2002). For increasing students' self-concept, building a less competitive learning environment to enhance students' self-concept in the classroom (Liem et al., 2013; Liou, 2014a, 2014b; Marsh & Craven, 2002) or to encourage students to pursue their personal academic goals instead of focusing on competing with others (Aschbacher et al., 2010; Carlone & Johnson, 2007) are also recommended.

Several weaknesses of this study and suggestions for future studies need to be addressed. First, the study did not take negatively worded items (e.g., science is boring) into consideration. Several studies (Marsh, 1996; Yang, Chen, Lo, & Turner, 2012) claim that including both positive and negative statements in a scale may yield method or artifactual factors. The multitrait-multimethod approach may be applied to further validate the factor structure in future studies (Byrne, 2011; Yang, Chen, Lo, & Turner, 2012). Second, this study only utilized data from one country as an exemplar. Although the results of this study have crucial implications for how to appropriately evaluate the optimal factor structure from the existing items to create derived variables, the three factor model composed of the 12 items from Taiwan could not be generalized to other countries. More studies related to analyzing data from other countries are needed to compare with the results of this study.

Conclusions

While ILSA provides so many data for researchers to use in their studies, appropriate quantitative methodologies should be utilized to examine items and to create optimal derived variables for research purposes. This study has identified and discussed three major methodological deficiencies that researchers should be aware of when utilizing data from ILSA. In the existing literature, the three common methodological issues are 1) the use of a single item rather than several items to define a latent variable, 2) the use of ordinal derived variables rather than interval derived variables for the subsequent statistical analyses, and 3) the use of the default factor structure rather than the optimal factor structure in a given situation. Moreover, due to the importance of students' attitudes toward learning science, it is essential to provide solid evidence to validate the use of such derived variables in a given situation.

Therefore, to overcome these statistical methodological issues, this study utilized an integrated approach to evaluating student attitudinal items and constructing derived variables in the TIMSS 2007 data for the Taiwanese eighth grade students. The results show that the optimal factor structure for the Taiwanese data is not the same as the default one. The results of this investigation should generate further discussion of rigorous analyses of survey items regarding attitudes towards learning science using ILSA data, not only for science education researchers in Taiwan but also internationally.

Acknowledgements

The author sincerely appreciates Dr. Frances Lawrenz and Dr. Ernest Davenport's feedback on the earlier version of this study. This study was supported by the National Science Council, Taiwan, under grant number 101-2410-H-008-049-.

References

Adams, R., J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21* (1), 1-23.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43* (4), 561-573.

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching, 47* (5), 564-582.

Bode, R. K. (1995, April). *Using Rasch to create measures from survey data.* Symposium conducted at the meeting of the American Educational Research Association, San Francisco, CA.

ISSN 1648–3898

Byrne, B. M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* Philadelphia, PA: Taylor and Francis.

Byrne, B. M., & Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Educational Psychology, 78*, 474-481.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44* (8), 1187-1218.

Chang, C. Y. (2005). Taiwanese science and life technology curriculum standards and earth systems education. *International Journal of Science Education, 27* (5), 625-638.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14* (3), 464-504.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding the using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14* (20). Available online: http://pareonline.net/pdf/v14n20.pdf.

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Eds.), *Achievement and achievement motive* (pp.76-146). San Francisco: W. H. Freeman.

Eklöf, H. (2007). Self-concept and valuing of mathematics in TIMSS 2003: Scale structure and relation to performance in a Swedish setting. *Scandinavian Journal of Educational Research, 51* (3), 297-313.

Foy, P., Galia, J., & Li, I. (2008). Scaling the data from the TIMSS 2007 and science assessments. In J. F. Olson, M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 225-379). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research, 71* (1), 105-131.

IDB Analyzer (Version 2) [Computer software and manual]. (2009). Hamburg, Germany: International Association for the Evaluation of Educational Achievement.

Kaya, S., & Rice, D. C. (2010). Multilevel effects of student and classroom factors on elementary science achievement in five countries. *International Journal of Science Education, 32* (10), 1337-1363.

Liem, G. A. D., Marsh, H. W., Martin, A. J., McInerney, D. M., & Yeung, A. S. (2013). The big-fish-little-pond effect and a national policy of within-school ability streaming: Alternative frames of reference. *American Educational Research Journal, 50* (2), 326-370.

Liou, P. Y. (2014a). Examining the big-fish-little-pond effect on students' self-concept of learning science in Taiwan based on the TIMSS databases. *International Journal of Science Education, 36* (12), 2009-2028.

Liou, P. Y. (2014b). Investigation of the big-fish-little-pond effect on students' self-concept of learning mathematics and science in Taiwan: Results from TIMSS 2011. *Asia-Pacific Education Researcher, 23* (3), 769-778.

Liou, P. Y., & Hung, Y. C. (in press). Statistical techniques utilized in analyzing PISA and TIMSS databases in science education from 1996 to 2013: A methodological review. *International Journal of Science and Mathematics Education.* DOI:10.1007/s10763-014-9558-5.

Liu, X., & Boone, W. (Eds.). (2006). *Applications of Rasch measurement in science education*. Maple Grove, MN: JAM Press.

Liu, S., & Meng, L. (2010). Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics self-concept. *Educational Psychology, 30* (6), 699-712.

Marsh, H. W. (1990). *Self Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph.* San Antonio, TX: Psychological Corporation.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70* (4), 810-819.

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker. P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology, 105* (1), 108-128.

Marsh, H. W., & Craven, R. G. (2002). The pivotal role of frames of reference in academic self-concept information: The big-fish-little-pond effect. In F. Pajares & T. Urdan (Eds.). *Adolescence and education* (Vol. II, pp. 83-123). Greenwich, CT: Information Age.

Marsh, H. W., Kong, C. K., & Hau, K. T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology, 93* (3), 543-553.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20* (3), 107-123.

Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the eighth and fourth grades*. Chestnut Hill, MA: Boston College.

Martin, M. O., & Preuschoff, C. (2008). Creating the TIMSS 2007 background indices. In J. F. Olson, M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 281-338). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muthén, L. K., & Muthén, B. O. (1998-2010). *MPlus (Version, 5.2)*. [Computer Software]. Los Angeles, CA: Muthén & Muthén.

Oliver, J. S., & Simpson, R. D. (1988). Influences of attitude toward science, achievement motivation, and science self concept on achievement in science: A longitudinal study. *Science Education, 72*, 143-155.

Olsen, R. V., Prenzel, M., & Martin, R. (2011). Interest in science: A many-faceted picture painted by data from the OECD PISA study. *International Journal of Science Education, 33* (1), 1-6.

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25* (9), 1049–1079.

Plucker, J., & Stocking, V. (2001). *Looking outside* and inside: Self-concept development of gifted adolescents. *Exceptional Children, 67*, 535-548.

Rinn, A. N., McQueen, K. S., Clark, G. L., & Rumsey, J. L. (2008). Gender differences in gifted adolescent's math/verbal self-concepts and math/verbal achievement: Implications for the STEM fields. *Journal for the Education of the Gifted, 32* (1), 34-53.

Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. M. Fayers & R. D. Hays (Eds.), *Assessing quality of life in clinical trials: Methods and practice* (2nd ed.). New York: Oxford University Press.

Sabah, S., Hammouri, H., & Akour, M. (2013). Validation of scale of attitudes toward science across countries using Rasch model: Findings from TIMSS. *Journal of Baltic Science Education, 12* (5), 692-702.

Schunk, D. H., Pintrich, P. R., & Meece, J., L. (2008). *Motivation in education* (3rd ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research, 95* (6), 323-332.

Shen, C., & Tam, H. P. (2008). The paradoxical relationship between student achievement and self-perception: a cross-national analysis based on three waves of TIMSS data. *Educational Research and Evaluation, 14* (1), 87-100.

SPSS (2008). *SPSS for Window, Rel. 17* [Computer Software]. Chicago IL: SPSS.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39* (2), 111-133.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89* (3), 451-469.

Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education, 72* (4), 331-346.

Wright, B. D., & Master, G. N. (1982). *Rasch scale analysis: Rasch measurement*. Chicago: MESA Press.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8* (3), 370.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). *ConQuest version 2.0: Generalized item response modeling software* [Computer software]. Melbourne, Australia: ACER.

Yang, Y., Chen, Y.-H., Lo, W.-J., & Turner, J. E. (2012). Cross-cultural evaluation as item wording effects on an attitudinal scale. *Journal of Psychoeduational Assessment, 30* (5), 509-519.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17* (2), 253-269.

Received: *July 28, 2014* Accepted: *October 30, 2014*

Pey-Yan Liou	Ph.D., Assistant Professor, Graduate Institute of Learning and Instruction & Center of Teacher Education, National Central University, No, 300, Jhongda Rd., Jhongli City, Taoyuan, 32001 Taiwan (R.O.C.). Tel: 886-3-4227151 ext. 26505 Fax: 886-3-4273371. E-mail: lioupeyyan@gmail.com