VALIDATION OF A SCALE OF ATTITUDES TOWARD SCIENCE ACROSS COUNTRIES USING RASCH MODEL: FINDINGS FROM TIMSS

Abstract. *The purpose of the present study was to validate a scale of attitudes toward science based on items and data of TIMSS 2007. The present research introduces several implications and recommendations for science education researchers who are interested in validating and using instruments to measure attitudes toward science. This research examined, from Rasch measurement perspectives, to what extent the scale was valid across diverse participating countries from Rasch measurement perspectives. The data of the highest three achieving countries (Singapore, Taiwan, and Japan), the lowest three achieving countries (Qatar, Botswana, and Ghana), and three average achieving countries (Australia, Scotland, and Italy) were analyzed using WINSTEPS software. Several Rasch analyses (e.g, fit statistics, PCA, and person/item map reliability analysis) were performed to validate the scale. Although most of the item fit the Rasch model, a few items were found to be misfitting. Then, the misfitting items were eliminated and a second analysis for the fitting eight items was performed to check on the validity of the scale. The results provided evidences that support the validity of the scale based on Rasch measurement.*

Key words: attitudes toward science, Rasch measurement, science education, validity.

Saed Sabah,
Hind Hammouri,
Mutasem Akour

Saed Sabah,
Hind Hammouri,
Mutasem Akour
Hashemite University, Zarqa, Jordan

Introduction

Attitudes toward science play an important role in science education (Hong & Lin 2011; Osborne, Simon, & Collins, 2003). Several international studies (e.g., House, 2008) found that students' attitudes toward sciences were a significant predictor of science achievement. Students' attitudes toward science in primary school are a major determinant of their responses to attitudes toward science at high schools (Murphy, Ambusaidi, & Beggs, 2006). George (2000) found that student' attitudes toward science decline over the years of middle and high school. Murphy, Ambusaidi, and Beggs (2006) found an evidence of decline in attitudes toward science as students get older . Osborne et al. (2003) reported that there had been a great decline in percentages of students studying science in UK. Preparing qualified scientists and engineers is required to develop a competitive workforce and enhance the economic prosperity (Osborne et al., 2003). Negative attitudes toward science may lead students to avoid science courses at high school and to move away from science for their future careers. Therefore, attitudes toward science should be monitored and assessed regularly taking into account that attitudes toward science are also an important dependent variable of curriculum evaluation.

The concept of attitudes toward science is not well understood (Osborne et al., 2003). As to Cheung (2009), this construct refers to feelings, beliefs, and values held by a person about science. Osborne et al. (2003) pointed out that attitudes toward science consists of a number of interrelated sub-constructs such as perceptions of science teacher, anxiety toward science, the value of science, self-esteem related to science, motivation toward science, enjoyment of science, attitudes of peers and friends toward science, attitudes of parents toward science, and the nature of the

Journal of Baltic Science Education, Vol. 12, No. 5, 2013

ISSN 1648–3898   VALIDATION OF A SCALE OF ATTITUDES TOWARD SCIENCE ACROSS COUNTRIES USING
RASCH MODEL: FINDINGS FROM TIMSS
(P. 692-702)

classroom environment. Therefore, the construct of attitudes toward science is a complex one that is constructed from several interrelated sub-constructs.

Measuring Attitudes toward Science

Developing a valid measure of attitudes toward science is necessary (Hong & Lin, 2011). Unfortunately, many of the developers of attitudes instruments failed to address the validity properly or even to provide a clear definition of attitudes toward science (Wang & Berlin, 2010). Cheung (2009) pointed out that many of attitude instruments suffered from lack of theoretical rationale and of empirical evidences to support construct validity. As to Wang and Berlin (2010), the development of attitude instrument should basically focus on three dimensions of attitudes toward science: (1) science enjoyment, (2) science confidence, and (3) importance of science. Based on the literature of more than three decades, Cheung (2009) developed a theoretical framework of attitudes toward chemistry which included four dimensions: (1) Liking for chemistry theory lessons (people have attitude when they love or hate science), (2) Liking for chemistry laboratory work, (3) Evaluative beliefs about school chemistry; and (4) Behavioral tendencies to learn chemistry. This framework adapted the idea that attitude is not a behavior but it is an action tendency to respond. In other words, students have positive attitudes toward science when they are willing to spend more time reading science books. Results of the confirmatory factor analysis supported Cheung's framework (see Cheung (2009) for more details). Researchers however usually overemphasize the first dimension and ignore the other dimensions such as liking laboratory work.

Attitudes scales are usually developed traditionally and attitude data which collected by Likert scale are inaccurately analyzed (Bond & Fox, 2007). Unfortunately, educational researchers routinely analyze ordinal-level attitudinal data as if they are equal interval (Boone, et al., 2010). They usually make mistakes when they directly carry out inferential statistics on ordinal data from attitude survey because inferential statistics assume data to be interval (Harwell & Gatti, 2001; Liu & Boone, 2006). Many of them also rely on Classical Test Theory (CTT) in developing and validating attitude instruments. However, Boone et al. (2010) indicated that Rach analysis could be used in preparing ordinal-level data for parametric analyses, and in developing higher quality science education instruments. Nam et al. (2011) suggested the use of Rasch model to overcome the limitations of CTT. Although Rasch analysis is basically quantitative, it is also 'richly qualitative' (Boone et al., 2010); therefore, it could serve as informative guidance while developing and validating instruments.

Trend in International Mathematics and Science Study (TIMSS 2007) developed three indices to measure eighth graders' attitudes toward science. These indices represent three aspects of students' attitudes: Positive affect, self confidence, and valuing science (Martin & Preuschoff, 2008). Confirmatory factor analysis was employed to examine the validity and dimensionality of the scales that were considered valid across countries (Martin & Preuschoff, 2008). Accordingly, the question is raised to investigate the extent to which TIMSS 2007 attitudes is a valid instrument across countries. Validating instruments across countries is important for developing fair large scale assessment for all students (Britton & Schneider, 2007; Nam et al., 2011)

The present study used the data of and items developed by TIMSS 2007 to validate an attitude scale across countries from the perspective of Rasch model.

Rasch Measurement Perspective

Using Rasch measurement model in developing and validating instruments has interest from educational researchers (e.g., Beglar, 2010; Cheng et al., 2011; Martinelli et al., 2009; Nam et al., 2011). The Use of Rasch model which produces interval measure makes the validation process more accurate (Liu & Boone, 2006). Many science education studies (e.g., Bradley & Sampson, 2006; Libarkin & Anderson, 2006) found that Rasch analysis was an effective tool in evaluating and improving the quality of science instruments by examining the validity of instruments in a non-traditional way. Rasch analysis also provides guidance for assessing the reliability and validity (content, construct, and predictive) of survey instruments (Boone et al., 2010).

Rasch model provides FIT statistics for items and persons to indicate how well each item or person fit the model. The FIT statistics are calculated based on the residuals between the observed and predicted responses (Liu & Boone, 2006). Rasch analysis programs (e.g., WINSTEPS,) provide INFIT and OUTFIT mean square (MNSQ) statistics to evaluate how well each item fits the model. MNSQ is a chi-square data-fit statistics based on the difference between the observed response patterns and the predicted response patterns; the closer MNSQ is to 1 the better" the fit (Liu & Boone, 2006). The deviations in this statistic from 1.0 refers to how much the data are less or more predictable than the model expects (Wright & Linacre, 1994). Bond and Fox (2007) consider fit as "a quality control principle" to determine whether items performances are meeting the requirements of Rasch model especially the unidimensionality requirement. Thus, fit statistics determine whether each item contribute to one construct only and whether the output could be interpreted on interval level measure (Bond & Fox, 2007). Linacre (2011) mentioned that Rasch model "analyzes the data as though they are unidimensional, and then the fit statistics report how well the data match the mathematically unidimensional framework that the Rasch analysis has constructed" (p. 1310). Bond and Fox considered INFIT and OUTFIT (MNSQ) values between 0.7 and 1.3 as acceptable indicators for fitting the model. As for the rating scale surveys, the acceptable range of MNSQ is 0.6 – 1.4 (Wright & Linacre, 1994).

Also, most Rasch software produce item-person map (Wright map) in which the estimations of abilities and difficulties are presented on the same continuum. The order and spacing of items on the hierarchical scale which are produced by Rasch analysis serve as a guide for improving instruments. As for Boone and Scantlebury (2006), the gaps on the hierarchical scale indicate a need for adding more items to fill these gaps and improve the instrument. Checking the order of item difficulty measures and comparing it with the theoretically hypothesized order may help identify and improve some items. To develop good instrument, it is necessary to develop items varying in their difficulties that target the attitude levels of respondents.

Furthermore, Rasch measurement provides two reliability indices: Item separation index (replicability of item placements) and person separation index (replicability of person ordering). These indices determine whether there are enough spread of items along the continuum and enough spread of ability among persons (Bond & Fox, 2007). Rasch reliability indicators are more accurate than the traditional (e.g., Cronbach's alpha) reliability indices as Rasch measurement estimates Standard Error of Measurement (SEM) for every person and for every item instead of assuming that SEM is the same for every person (Bond & Fox, 2007; Liu & Boone, 2006; Smith, 2004). The traditional reliability indices usually include the extreme scores and overestimate the internal consistency of instruments (Linacre, 1997).

Rasch theory provides a framework as well as a road map for survey design and analysis (Boone et al., 2010). Researchers (e.g., Cheng et al., 2011; Lee- Ellis, 2009) usually focus on fit statistic to check on the quality of items by investigating how well they fit the Rasch unidimensional model. Fitting the Rasch model provides evidence for the construct validity; it could be considered as an indication that the hypothesized construct exists (Liu & Boone, 2006). Using Rasch model in developing instruments is a theory-based approach; when there is a good model-data-fit, then there is an evidence for the construct validity of measures (Liu, 2006). The present research used the Rasch measurement to validate attitude scale across countries based on participants' responses to attitudes items that were developed by TIMSS (2007). Developing a valid attitude scale across countries may help researchers in collecting attitudinal data accurately and conducting comparative studies. Also, the current research may encourage educational researchers to pay more attention to the cross-countries validation of other instruments that could be used in the large scale comparative studies.

Methodology of the Study

Background on Attitudes toward Science Indices Developed by TIMSS 2007

TIMSS study developed items which reflect three aspects of students' attitudes toward science: Positive Affect toward Science (PATS), Self-Confidence in learning Science (SCS), and Students' Valuing Science (SVS) (Martin & Preuschoff, 2008). Students' responses were based on a 4-point scale: *agree a*

Journal of Baltic Science Education, Vol. 12, No. 5, 2013

ISSN 1648–3898   VALIDATION OF A SCALE OF ATTITUDES TOWARD SCIENCE ACROSS COUNTRIES USING
RASCH MODEL: FINDINGS FROM TIMSS
(P. 692-702)

lot = 1, *agree a little* = 2, *disagree a little* = 3, and *disagree a lot* = 4. (PATS) is based on students' responses to three items: (1) I enjoy learning science; (2) Science is boring; and (3) I like Science. SCS is based on students' responses to four items: (1) I usually do well in science; (2) I learn things quickly in science; (3) Science is more difficult for me than for many of my classmates; and (4) Science is not one of my strengths. The third aspecte, SVM, is composed of three statements: (1) I think learning science will help me in my daily life; (2) I need science to learn other school subjects; and (3)

I need to do well in science to get into the university of my choice; and (4) I would like to do well to get the job I want. Before analyzing the data, the negative items were reverse coded.

Selected Data and Data Analysis

Because the purpose of the present research was validating the attitudes toward science scale across countries, the data of nine countries were selected. The selected countries were the highest three achieving countries (Singapore, Taiwan, and Japan), the lowest three achieving countries (Qatar, Botswana, and Ghana), and three average achieving countries (Australia, Scotland, and Italy). A random sample of 150 participants was selected from the data of each country. The three negatively worded statements were reverse coded before conducting the analysis. WINSTEPS software (Linacre, 2005) was used to analyze the data. Reliability analysis, fit statistics, person/item map, and Principal Component Analysis (PCA) were used to validate the scale.

Procedures

First, the fit statistics for the eleven attitudes items (stage 1) was conducted. The results revealed that the negatively worded items, 3 items, usually outfitted the acceptable range of fit statistics for most of the courtiers. After that, the misfitting items were eliminated and a second stage analyses (fit statistics, reliability analysis, and person/item map, PCA) were performed to check on the quality of items and the validity of the construct (Linacre, 2010).

Results of the Study

Fit Statistics

Rasch analysis estimates the difficulty of each item which is, in case of a Likert scale, the level of "agreement with" or "endorsement of" (Bond & Fox, 2007). The item fit helps educators to identify to what extent each of the eleven items elicits unexpected responses on the part of respondents. Fit statistics could be used for evaluating the quality of each item (Bond & Fox, 2007). The acceptable range of infit and outfit MNSQ is 0.60 to 1.4 (Wright & Linacre, 1994). The Infit and Outfit MNSQ of the eleven items across the nine countries were investigated. In general, three items were found to outfit the acceptable range of MNSQ for one or more countries (Table 1).

To give more details, the data of Singapore indicated that the values of MNSQ of one item (science is more difficult for me) outfitted the acceptable range of MNSQ. As for the data of Taiwan, three items (science is more difficult for me than for many of my classmates; science is not one of my strength; and I need to do well in science to get into the university of my choice) outfitted the acceptable range of MNSQ. The data of Japan indicated that two items outfitted the model. These items are "science is more difficult for me than for many of my classmates", and "I need to do well in science to get the job I want". To summarize, the item "science is more difficult for me" outfitted the acceptable range of MNSQ for the highest three achieving countries.

Table 1. Infit and Outfit MNSQ of the three misfitting items (Stage 1).

Item Statement	Highest Achieving Countries						Average Achieving Countries						Lowest Achieving Countries					
	Singapore		Taiwan		Japan		Australia		Scotland		Italy		Qatar		Ghana		Botswana	
	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ
Science is boring	-.37	1.18 1.10	-0.35	0.78 0.76	-.75	1.23 1.23	.23	1.11 1.12	.37	1.31 1.14	-.45	1.35 1.74	.29	1.43 1.24	.63	2.18 3.83	-.05	1.07 .95
Science is more difficult for me than for many of my classmates	0.70	2.3 -1.9	0.04	1.37 1.58	-.17	1.40 1.51	-.43	1.93 2.29	-.13	1.49 1.63	-.50	.72 .85	.69	1.52 1.75	.95	1.69 1.72	.54	1.40 1.27
Science is not one of my strengths	1.23	0.98 0.98	0.35	1.64 1.89	.19	1.16 1.16	.47	1.36 1.61	.06	1.23 1.02	.05	1.7 1.48	.11	2.22 3.00	.53	1.75 2.40	.26	2.28 2.78

Table 2. Infit and Outfit MNSQ of the 8 items (Stage 1 & Stage 2).

Item Statement	Highest Achieving Countries						Average Achieving Countries						Lowest Achieving Countries					
	Singapore		Taiwan		Japan		Australia		Scotland		Italy		Qatar		Ghana		Botswana	
	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ
I enjoy learning science.	-0.63	0.79 0.72	-.06	0.77 0.75	-.46	0.70 0.70	-.10	.80 .80	-.02	.72 .70	-.04	.76 .76	-.11	.66 .59	-.31	.75 .78	-.43	.68 .60
I enjoy learning science *	-0.5	0.91 0.83	-.07	0.93 0.88	-.68	0.88 0.89	-.14	.91 .90	.05	.84 .82	-.07	1.08 1.03	.09	.82 .71	-.07	.80 .67	-.39	1.25 1.09
I need to do well in science to get the job I want	-0.23	1.4 1.29	0.27	1.13 1.13	.11	1.37 1.45	.16	1.22 1.02	.16	1.32 1.13	.75	1.18 1.18	-.21	1.00 .94	-.48	0.45 .58	-.09	1.21 0.90
I need to do well in science to get the job I want*	-0.04	1.29 1.15	0.33	1.11 1.09	.01	1.36 1.40	.76	1.10 1.08	.24	1.26 1.15	.19	1.15 1.08	-.15	1.01 .77	-.38	1.30 1.12	.12	.98 .57
I think learning science will help me in my daily life	-0.76	0.87 0.95	-1.04	1.07 1.16	-.36	.76 0.75	-.23	.65 .72	-.24	0.80 .77	-.49	.81 .79	-.50	.48 .51	-.52	.96 2.70	-.33	1.03 0.98
I think learning science will help me in my daily life *	-0.64	0.88 0.87	-1.25	1.12 1.28	-.55	.72 .72	-.79	.80 .81	-.30	0.90 .97	-.25	.77 .82	-.62	.94 .76	-1.02	1.18 1.03	-.51	1.00 .75
I like science	-0.67	0.92 0.87	0.02	0.61 0.61	-.23	.68 .67	.09	.95 .89	.18	.74 .69	.60	.93 .91	.16	.61 .53	-.37	.62 .69	-.52	.58 .47
I like science *	-0.55	1.07 0.96	0.03	0.75 0.72	-.40	.84 .83	.58	1.1 1.07	.29	.92 .89	.19	1.21 1.17	.51	.74 .64	-.06	1.18 .84	-.77	1.1 .64

Journal of Baltic Science Education, Vol. 12, No. 5, 2013

ISSN 1648–3898   VALIDATION OF A SCALE OF ATTITUDES TOWARD SCIENCE ACROSS COUNTRIES USING RASCH MODEL: FINDINGS FROM TIMSS
(P. 692-702)

| Item Statement | Highest Achieving Countries | | | | | | Average Achieving Countries | | | | | | Lowest Achieving Countries | | | | | |
| | Singapore | | Taiwan | | Japan | | Australia | | Scotland | | Italy | | Qatar | | Ghana | | Botswana | |
	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ	Mes.	Infit Oufit MNSQ
I need science to learn other school subjects	0.04	1.13 1.35	0.32	0.93 0.93	.88	.93 .97	.12	.86 .95	.19	.92 1.02	.28	1.04 1.08	-.08	.72 .66	-.18	.59 .68	.21	1.06 .90
I need science to learn other school subjects*	0.27	1.09 1.14	0.39	0.99 0.95	.98	.97 .94	.21	1.04 1.17	.27	.94 1.03	.22	0.92 0.98	.10	.80 .78	.35	1.00 1.24	.48	1.12 1.10
I learn things quickly in science	0.79	0.68 0.73	0.56	0.82 0.81	.55	.72 .75	.02	.73 .72	.02	.76 .74	-.30	.58 .66	.42	.68 .70	.19	.70 .81	.40	.74 0.77
I learn things quickly in science *	1.12	0.80 0.84	0.70	1.03 0.99	.56	.91 .93	-.45	.78 .80	.10	.98 .96	.08	1.02 1.09	.89	1.02 1.19	1.01	.83 0.70	.80	.92 .94
I need to do well in science to get into the university of my choice	-0.78	1.1 1.03	-0.07	1.28 2.4	-.41	1.30 1.27	.03	.82 .82	-.19	1.10 1.04	.42	1.23 1.24	-.34	.53 .49	-.40	.73 4.52	-.17	.87 .77
I need to do well in science to get into the university of my choice *	-0.67	1.05 0.94	-0.09	1.29 1.22	-.61	1.25 1.25	.39	1.19 1.19	-.22	1.06 1.03	.10	0.90 0.90	-.31	.68 .68	-.31	1.22 1.21	-.02	1.14 .91
I usually do well in science	0.68	0.72 0.73	-0.04	0.53 0.56	.65	.70 .76	-.37	.61 .72	-.39	.75 .80	-.33	.75 .75	-.45	1.19 2.00	-.05	.95 .79	.19	.84 .75
I usually do well in science *	1.00	1.02 1.02	-0.05	0.74 0.81	.69	.98 1.03	-.56	.93 1.05	-.42	1.04 1.08	-.46	.88 0.91	-.51	2.39 3.01	.47	1.08 1.18	.42	1.09 1.17

As for the average achieving countries, fit statistics revealed that few items outfitted the acceptable range of MNSQ. As for the data of Australia, the items that outfitted the acceptable range were "science is more difficult for me than for many of my classmates" and "science is not one of my strength". For Scotland, one item outfitted the acceptable range: "science is more difficult for me than for many of my classmates". The data of Italian students revealed that two items outfitted the model. These two items are "science is boring" and "science is not one of my strength". Therefore, these two items could be considered problematic for the average achieving countries.

As for the lowest achieving countries, the number of outfitting items was larger than those of average- or highest- achieving countries. Four items from Qatar data and five items from Ghana outfitted the acceptable range of MNSQ. As for the results of Qatar and Ghana, the three negatively worded items outfitted the acceptable range of MNSQ. These items were "science is not one of my strength", "science is boring", and "science is more difficult for me than for many of my classmates". Accordingly, we decided to remove the three misfitting items and perform a second series of analyses (stage 2).

Fit Statistics (Stage 2)

After eliminating the negatively worded items, the infit and outfit statistics for the items across countries were within the acceptable range of MNSQ with a few exceptions as shown in Table 2. This basically means that the data of the selected countries met the basic assumptions of Rasch model. Fitting the Rasch model could be considered as evidence that support the construct validity (Liu & Boone, 2006) of the eight-item scale.

Wright Maps: Item/ Person Maps

Investigating the maps of the high achieving countries (Figure 1) indicated that the items targeted the levels of attitudes of participants. The estimates of "item difficulty to endorse" ranged from -1 logit to +1 logit. However, the match between item difficulties and person abilities was not perfect especially for Singapore and Taiwan. In the case of Singapore and Taiwan, there were no items to target the participants of high level of attitudes, above + 1 logit. Further, there were no items to target many Taiwanese whose levels of attitudes were below -1 logit. This means that the levels of attitudes of persons whose levels were above +1 logit or below -1 logit were estimated imprecisely (Bond & Fox, 2007). However the map of Japan was the best among the nine maps we have investigated in the present research. Although the agreement levels for the items ranged from -1 logit to +1 logit, the levels of attitudes for the Japanese approximately ranged from -1 logit to +1 logit, which indicated a good match between "item difficulty to endorse" and students' levels of attitudes.

Figure 1: **Wright maps of the high achieving countries for 11 items of Attitude toward Science Scale.**
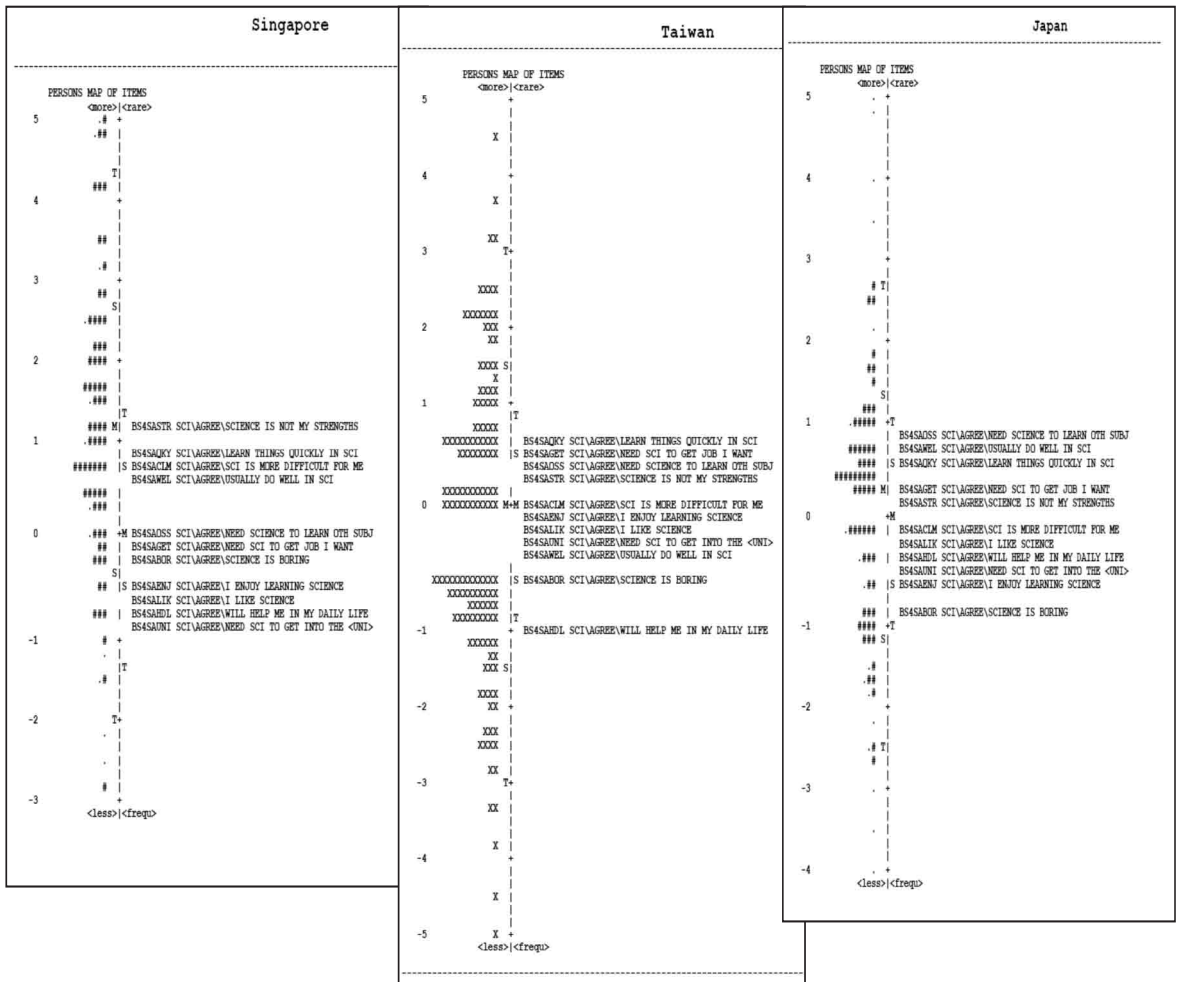
The maps of the average achieving countries were also investigated. The items were estimated to be very difficult 'to agree with' for the participants of the three average achieving countries, Australia, Scotland, and Italy; the endorsement levels were approximately ranged from -0.50 logit to 0.50 logit. It was noticeable that no item was targeting many participants whose level of attitudes was below -1 logit.

Journal of Baltic Science Education, Vol. 12, No. 5, 2013

ISSN 1648–3898   VALIDATION OF A SCALE OF ATTITUDES TOWARD SCIENCE ACROSS COUNTRIES USING
RASCH MODEL: FINDINGS FROM TIMSS
(P. 692-702)

Also the maps for the three lowest achieving countries, Qatar, Botswana, and Ghana were investigated. With regard to Qatar and Ghana, the difficulty levels of the items ranged approximately from -0.6 to 1.0 logit. The items did not very well target the students' levels of attitudes that the items appeared very difficult to agree with to many low- attitude students from Qatar and Ghana. However, the students of Botswana expressed high level of attitudes compared to their counterparts in the other low achieving countries. Therefore, the items failed to target the high –attitude students whose attitude level were above 1.0 logit. In sum, although there was acceptable match between the levels of endorsement and attitudes, especially for those of high-achieving countries, the maps informed us that the items, especially those of lowest-achieving, did not perfectly target the participants' levels of attitudes, which might lead to inaccurate estimates.

Principal Component Analysis (PCA)

To collect more evidences that support the unidimensionality of construct, The PCA was conducted using Winsteps software for the two formats of the scale. The results of PCA show that the percentage of variance explained by the measures (11 items) was strong (> 40 %) for the average and high achieving countries (Table 3). It was low for some low achieving countries. However, for stage 2, the variance explained by the measures was high (> 56 %) for all the participating countries which support the dimensionality of the construct. Another evidence that supported the dimensionality of the construct was the low percentage of unexplained variance by the first factor as shown in Table 3. For example, the strength of Rasch dimension for the data of Taiwan was 68 % which could be compared with the strength of the biggest secondary dimension (11.1%) indicating that the attitudes toward science could be considered unidimensional for practical purposes (Linacre, 2003). In short, the results of PCA provided evidence that supported the construct validity of the scale.

Reliability Analysis

As for stage 1, the person reliability indices ranged from 0.64 to 0.89 (Table 4) for the selected nine countries. The indices of the highest achieving countries (0.81 – 0.86) and the indices of the average achieving countries (0.86 – 0.89) were higher than those of the lowest achieving countries (0.64 – 0.77). The person reliability indices for the average- and the highest achieving countries were relatively high (> = 0.81).

The item reliability indices ranged from 0.67 to 0.96 as shown in Table 4. The item reliability indices were generally high (> = 0.89) for the most of the selected countries. The high values of item reliability indices indicate that the item estimates to be replicated when the attitudes scale is applied to other samples or countries. High reliability indices also suggest that the instrument differentiates the levels of attitudes successfully.

Table 3. **Standardized Residual Variance of Principal Component Analysis for the Attitude toward Science Scale (stage 1 & stage 2).**

Country	Percentage of variance explained by measure		Percentage of variance unexplained (variance explained by 1st factor)	
	Stage 1 (11 items)	Stage 2 (8 items)*	Stage 1 (11 items)	Stage 2 (8 items)*
Singapore	52.9	56.5	12.2	14.3
Taiwan	65.2	68	9.3	11.1
Japan	61	68	9.1	10.3
Australia	59.8	65.5	11.3	13.3
Scotland	58.8	65.8	11.3	12.9

Country	Percentage of variance explained by measure		Percentage of variance unexplained (variance explained by 1st factor)	
	Stage 1 (11 items)	Stage 2 (8 items)*	Stage 1 (11 items)	Stage 2 (8 items)*
Italy	53.3	63.8	10.6	10.8
Qatar	44.9	68.8	12.7	8
Ghana	2	57	20	9.4
Botswana	27.7	56.8	12.9	9.5

The three misfitting items were excluded in stage 2

Table 4.   Person and Item reliability Indices (stage 1 and stage 2).

Item Reliability	Person Reliability	Lowest Achieving Countries	Item Reliability	Person Reliability	Average achieving Countries	Item Reliability	Person Reliability	High Achieving Countries
0.91 (0.91*)	0.77 (0.69*)	Qatar	0.77 (0.63*)	0.89 (0.85*)	Australia	0.96 (0.95*)	0.86 (0.84*)	Singapore
0.91 (0.93*)	0.72 (0.51*)	Botswana	0.67 (0.74*)	0.87 (0.83*)	Scotland	0.89 (0.93*)	0.85 (0.84*)	Taiwan
0.95 (0.90*)	0.64 (0.49*)	Ghana	0.91 (0.93*)	0.86 (0.82*)	Italy	0.92 (0.94*)	0.81 (0.80*)	Japan

Stage 2 , the negatively worded items were excluded

Reliability Analysis (Stage 2)

After eliminating the misfitting items, the person reliability indices for the data of both high achieving and average achieving countries were not largely affected, they are above .80 as shown in Table 3. On the other hand, in case of the low achieving countries, the indices decreased greatly especially for the data of Ghana where it reached .49.

As for the item reliability indices, they were above .90 for the data of seven countries even after eliminating the three misfitting items. Nevertheless, the item reliability for the data of Australia was relatively low (.63). In short, the item and person reliability indices were relatively high for most of the countries even after eliminating three items. The main concern was the low person reliability indices of the lowest achieving countries.

Discussion

The purpose of this research was to validate a scale of attitudes which could be used by science educators. The scale covers three aspects of attitudes: Positive Affect toward Science (PATS), Self-Confidence in learning Science (SCS), and Students' Valuing Science (SVS) (Martin & Preuschoff, 2008). This scale was constructed based on the main aspects of attitudes that were identified by science education researchers over the past three decades (Liu, 2010; Osborne et al., 2003). The item and person reliability indices were relatively high for most of the countries.

Several evidences that supported the validity of the scale were presented. First, the infit and outfit statistics for the items across countries were within the acceptable range of MNSQ with a few exceptions. Fitting the Rasch model could be considered as an evidence that supports the construct validity of the suggested scal; the eight-item scale. In other words, fitting the Rasch model provides evidence for the construct validity of the scale (Liu & Boone, 2006). Therefore, it could be concluded that the items

Journal of Baltic Science Education, Vol. 12, No. 5, 2013

ISSN 1648–3898   VALIDATION OF A SCALE OF ATTITUDES TOWARD SCIENCE ACROSS COUNTRIES USING
RASCH MODEL: FINDINGS FROM TIMSS
(P. 692-702)

are of high quality with regard to meeting the undimensionality requirement and contributing to one construct (Bond & Fox, 2007). Furthermore, the item –person map indicated that there was fair good spread of persons and items especially for the high achieving countries. The maps show that there was no redundancy of items. Also, the PCA provided evidence that support the construct validity. The results of PCA indicated that the variance explained by the measure was above 56 % for the nine selected countries which could be considered high (Conrad et al., 2009). The strength of the Rasch dimension provided evidence that support the construct validity of the scale (Linacre, 2003).

Implications and Suggestions for Further Research

The present study provided evidences that support the validity of the 8-item attitude scale across countries especially for the high- and average- achieving countries. This validated scale covers the main aspects of attitudes; therefore, educators are encouraged to use this scale to collect attitudinal data and conduct comparative studies. Continuing research on students' attitudes is needed to understand and remediate the decline in numbers of students studying science (Osborne et al., 2003). TIMSS and other international studies are advised to pay more attention to the cross-countries validation of scales using the Rasch model.

Psychometricians are also invited to investigate why the negatively worded items were misfitting. Interviewing students could be an excellent technique which would provide researchers with information about how students perceive and understand these statements. Moreover, TIMSS should be more sensitive to any factors that could distort the validity of the scale across countries.

The present research introduces some suggestion for future research. Since the validated scale did not include items of "Liking Laboratory work" (Cheung, 2009) and students' belief of relevance of science to making better decisions (Siegel & Ranney, 2003), researchers are invited to consider them when they develop new scales of attitudes toward science. The results of fit statistics across countries indicated that the negatively worded items were problematic. Therefore, further investigations are needed to explain why the negatively worded items were more likely to misfit Rasch the model.

References

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, *27*(1), 101-118.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates Publishers, New Jersey.

Boone, W. J., Townsend, J. S., & Staver, J. (2010). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education, 95* (2), 258-280.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90* (2), 253-269.

Bradley, K. D., & Sampson, S. O. (2006). Constructing a quality assessment through Rasch techniques: The process of measurement, feedback, reflection and change. In X. Liu & W. Boone (Eds.), *Applications of Rasch measurement in science education* (pp. 23-44). Maple Grove, MN: JAM Press.

Britton, E. D., & Schneider, S. A. (2007). Large-scale assessments in science education. In S. K Abell & N. G. Lederman (Eds.). *Handbook of Research on Science Education* (pp. 1007-1040), Mahwah, NJ: LEA.

Cheng, Y. Y., Chen, L. M., Liu, K. S., & Chen, Y. L. (2011). Development and psychometric evaluation of the school bullying scales: A Rasch measurement approach. *Educational and Psychological Measurement*, *71* (1), 200- 216.

Cheung, D. (2009). Developing a scale to measure students' attitudes toward chemistry lessons. *International Journal of Science Education, 31* (16), 2185-2203.

Conrad, K. J., Conrad, K. M., Dennis, M. L., Riley, B. B., Chan, Y. F., & Funk, R., (2009). *Validation of the crime and violence scale (CVS) to the Rasch measurement model, GAIN Methods Report 1.1.* Bloomington, IL: Chestnut Health Systems.

George, R. (2000). Measuring change in students' attitudes toward science over time: An application of latent variable growth modelling. *Journal of Science Education and Technology, 9* (3), 213- 225.

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

Hong, Z. R., & Lin, H. (2011). An investigation of students' personality traits and attitudes toward science. *International Journal of Science Education*, *33* (7), 1001-1028.

House, J. D. (2008). Effect of classroom instructional strategies and self-beliefs on science achievement of elementary-

school students in Japan: Results from the TIMSS 2003 assessment. *Education, 129* (2), 259-266.

Lee-Ellis, S. (2009). The development and validation of a Korean C-Test using Rasch analysis. *Language Testing, 26* (2), 245-274.

Libarkin, J. C., & Anderson, S. W. (2006). The Geoscience Concept Inventory: Application of Rasch Analysis to Concept Inventory Development in Higher Education. In X. Liu & W. Boone (Eds.), *Applications of Rasch measurement in science education* (pp. 45-73). Maple Grove, MN: JAM Press.

Linacre, J. M. (1997). KR-20 or Rasch reliability: Which tells the "Truth"? *Rasch Measurement Transactions, 11*(3), 580-601.

Linacre, J. M. (2003**).** Data variance: Explained, modeled and empirical. *Rasch Measurement Transactions*, *17*(3), 942-943.

Linacre, J. M. (2005). WINSTEPS Rasch measurement computer program. Chicago, IL:Winsteps.com

Linacre J.M. (2010) When to stop removing items and persons in research misfit analysis?, *Rasch Measurement Transactions, 23* (4), 1241.

Linacre, J. M. (2011). Rasch measures and unidimensionality. *Rasch Measurement Transactions, 24* (4), 1310.

Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch Modelling approach*. Charlotte, NC: Information Age Publishing

Liu X. and Boone W. J., (2006), Introduction to Rasch measurement in science education, in Liu, X. and Boone, W. J. (eds.), *Applications of Rasch measurement in science education* (pp. 1-22). Maple Grove, Minnesota, JAM Press.

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.

Martinelli, S., Bartholomeu, D., Caliatto, S. G., & Sassi, A. (2009). Children's self-efficacy scale: Initial psychometric studies. *Journal of Psychoeducational Assessment*, *27* (2), 145-156.

Murphy, C., Ambusaidi, A., & Beggs, A. (2006). Middle East meets West: Comparing children's attitudes to school science. *International Journal of Science Education, 28* (4), 405-422.

Nam, S. K., Yang, E., Lee, S., Lee, S. H., & Seol, H. (2011). A psychometric evaluation of the career decision self-efficacy scale with Korean students: A Rasch model approach. *Journal of Career Development*, *38* (2), 147-166.

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

Siegel, M. A., & Ranney, M. A. (2003). Developing the changes in attitude about the relevance of science (CARS) questionnaire and assessing two high school science classes. *Journal of Research in Science Teaching*, *40* (8), 757-775.

Smith, E. (2004). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. Smith & R. Smith (Eds.), An introduction to Rasch measurement (pp. 93–122). Maple Grove, MN: JAM Press.

Tang H. (1996) What is Rasch? *Rasch Measurement Transactions, 10* (2), 507.

Wang, T. L., & Berlin, D. (2010). Construction and validation of an instrument to measure Taiwanese elementary students' attitudes toward their science class. *International Journal of Science Education, 32* (18), 2413-2428.

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

Saed Sabah	PhD., Assistant Professor of Science Education, Department of Teaching and Curricula, Faculty of Educational Sciences, Hashemite University, Jordan. E-mail: saed_sabah@yahoo.com or Sabah@hu.edu.jo
Hind Hammouri	PhD., Professor of Measurement and Evaluation , Department of Educational Psychology, Faculty of Educational Sciences, Hashemite University, Jordan. E-mail: hind@hu.edu.jo
Mutasem Akour	PhD., Assistant Professor Measurement and Statistics , Department of Educational Psychology, Faculty of Educational Sciences, Hashemite University, Jordan. E-mail: mutasem@hu.edu.jo