Constructs Evaluation of Student Attitudes towards Science

Liu Huang [1], Fan Huang [2], Pey Tee Oon [2*], Miranda Chi Kuan Mak [2]

[1] Huiyang school affiliated to SCNU, Huiyang, Huizhou, Guangdong, CHINA
[2] University of Macau, CHINA

ABSTRACT

Students' attitudes towards science (SAS) is a prominent research area as evident in the science education literature. Many SAS studies incorporated science confidence (SC), science enjoyment (SE) and importance of science (IS) into the study of SAS. However, the incorporation of these constructs often depends on the subjective judgment of the researchers. This study examines the incorporation of the three constructs in the measure of SAS based on the Asian Student Attitudes Towards Science Class (ASATSC) instrument. A total of 1,133 7th to 11th graders from China completed a survey of the three constructs. Data was collectively assessed in terms of fit to the Rasch model that requires invariant and consistent response category functioning. Results indicated that SC was not correlated well with IS whilst SC and SE were consistent with each other in the measure of SAS. Recommendations were provided on how constructs on the measurement of SAS can be better designed.

Keywords: attitudes towards science, constructs, Rasch model, Chinese context, secondary school students

BACKGROUND

The study of SAS is a prominent research area. Pell and Jarvis (2001) highlighted that attitudes affect students' attention, consistence, and behavior in the classroom (Germann, 1988; Weinburgh, 1995). Students who perceived science positively are more likely to pursue this subject after the compulsory education (Pell & Jarvis, 2001). On the contrary, a negative image dissuades them from it (Trumper, 2006). As evident in the science education literature, many scholars grappled with students' negative image towards science (e.g., Anwer, Iqbal, & Harrison, 2012; Barmby, Kind, & Jones, 2008; Potvin & Hasni, 2014).

Theoretical Background

A problem regarding the measurement of SAS is that there is no consensus among researchers as to which constructs the SAS scale should included. In 1978, Fraser developed a Test of Science Related Attitudes (TOSRA) including constructs of 'Social Implications of Science,' 'Normality of Scientists,' 'Attitudes toward Scientific Inquiry,' 'Adoption of Scientific Attitudes,' 'Enjoyment of Science Lessons,' 'Leisure Interest in Science,' and 'Career Interest in Science.' Pell and Jarvis (2001) included scales of 'liking school,' 'independent investigator,' 'science enthusiasm,' the 'social context of science,' and 'science as a difficult subject' in measuring SAS. Murphy and Beggs (2003) proposed SAS as 'enjoyment of science,' 'appreciation of the importance of science,' and 'perceived ability to do science.' PISA 2006 evaluated attitudes towards science by including the 'interest in science,' 'support for scientific enquiry,' and 'responsibility for sustainable development' constructs (Bybee & McCrae, 2011). Wang and Berlin (2010) in Taiwan included constructs of 'science enjoyment,' 'science confidence,' and 'importance of science as related to science class experiences' in assessing SAS while Wan and Lee (2017) in Hong Kong advocated the constructs of 'behavior,' 'cognition,' and 'affect.' It seems hard to decide which constructs are most appropriate to be incorporated as the measurement of SAS is a complex issue, dependent on researchers' diverse interests to look at specific aspects in their studies, and further complicated by cultural considerations (Oon & Subramaniam, 2018).

Contribution of this paper to the literature

- Problem: Subjective use of student attitudes towards science (SAS) constructs in pertinent literature.
- Significance of current study: SAS constructs were meaningfully evaluated by Rasch Measurement Model that supports invariant structure of responses. Results are theoretically discussed based on Rasch measurement results.
- Findings: Two constructs commonly incorporated in the SAS survey instrument were found to be inconsistent with each other. Recommendations are provided on how constructs can be better designed and interpreted for the measurement of SAS.

The commonly used SAS constructs reported in the pertinent literature include confidence in science (self-efficacy/self-concept), student like learning science (enjoyment, interest, motivation), and value of science (confidence, abilities, belief) using Likert scales as used and reported in the TIMSS international study (2011, 2015) (Michael, Mullis, Foy, & Stanco, 2012; Mullis & Martin, 2013), and in many other independent studies, see below:

(a)   Confidence in science (e.g., Bryan, Glynn, & Kittleson, 2011; DeBacker & Nelson, 2000; Greene & Miller, 1996; Lee & Brophy, 1996, Nolen & Haladyna, 1989; Meece, Blumenfeld, & Hoyle, 1988; Miller, Greene, Montaivo, Ravindran, & Nichols, 1996; Nolen & Haladyna, 1990; Tuan, Chin, & Shieh, 2005).

(b)   Interest in and enjoyment of science (e.g., Bathgate, Schunn, & Correnti, 2014; Harackiewicz, Rozerk, Hulleman, & Hyde, 2012; Lee & Brophy, 1996; Meece, Anderman, & Anderman, 2006; Murphy & Beggs, 2003; Nolen & Haladyna, 1989; Osborne, Simon, & Collins, 2003; Palmer, 2005; Tuan, Chin, & Shieh, 2005; Zeyer, Bölsterli, Brovelli, & Odermatt, 2012).

(c)   Value/relevance of learning science (e.g., Bathgate, Schunn, & Correnti, 2014; Murphy & Beggs, 2003; Nolen & Haladyna, 1989; Tuan, Chin, & Shieh, 2005).

Confidence in science

In a study by Wang and Berlin (2010), science confidence was recognized as *'the extent to which a student is confident and feels successful in science class'* (*p*. 2418). Confidence was *relevant to motivational belief governed by students' belief of own ability* (Bryan, Glynn, & Kittleson, 2011, *p*. 1050; Simpkins, Price, & Garcia, 2015, *p*. 1387). Grades that students received in class, the success or failure towards science process, and evaluation of capabilities influenced students' motivation and confidence in the subject (Nolen & Haladyna, 1989, reported in Tuan, Chin, & Shieh, 2005, *p*.642; Sheldrake, 2016). The degrees of self-confidence have been found to be able to predict educational outcomes (Sheldrake, 2016).

Interest in and enjoyment of science

The students' enjoyment of science has often been referred to as intrinsic motivation, as '*doing something because it is inherently interesting or enjoyable*' (Ryan & Deci, 2000, *p*. 55 as reported in Palmer, 2005, *p*.1858). That means a person feels the pleasure or has fun when s/he is doing something s/he is naturally passionate about. Oon and Subramaniam (2013) considered it as internal factor and concluded that the difficult nature inherent in physics has often hindered the sense of pleasure to study science subjects after compulsory education, as reported in earlier studies (e.g., Kessels, Rau, & Hannover, 2006).

Relevance of learning science

George (2006) reported that the positive attitudes towards science are associated with positive attitudes towards the utility of science. According to Brophy (1998), motivation to learning is '*a tendency to find academic activities meaningful and worthwhile and to try to derive the intended academic benefits from them*' (Cavas, 2011, *p*.32). In PISA, the relevance of learning science was described *as 'general value of science'* (OECD, 2007) as in TIMSS 2011 and 2015 (Michael, Mullis, Foy, & Stanco, 2012; Mullis & Martin, 2013). Encouragement from teachers and their attention on the teaching of science can foster positive attitudes towards the utility of science (George, 2006).

Rationale for the Current Study

Methodological and conceptualization issues of SAS constructs measurement

Most SAS scales use Likert rating scales comprising various subscales in measuring different aspects of SAS (Angell, Guttersrud, Henriksen, & Isnes, 2004; Pell & Jarvis, 2001; Stokking, 2000; Wang & Berlin, 2010). A few recent SAS studies in China reported using Likert scales (e.g., Li & Chen, 2016; Li & Yao, 2014; Wan & Lee, 2017).

Issues of validity and reliability stemming from the data using Likert scales in these studies remained questionable. The linearity of the Likert scales is often wrongly assumed. In fact, there is no guarantee of definite equidistance between response categories (e.g., distance between *Strongly Disagree* with Disagree and *Strongly Agree* with *Agree*) (Wright, 1996). The raw scores from the rating scale are not linear, which may compromise the reliability and validity of results (Wright, 1996; Wright & Masters, 1982). Therefore, we used the Rasch Model analysis in the present study to overcome this psychometric issue.

As mentioned above, the common SAS constructs accepted by researchers consist of '*science enjoyment*' (SE), '*science confidence*' (SC) and '*importance of science*' (IS) (e.g., Bathgate, Schunn, & Correnti, 2014; Bryan, Glynn, & Kittleson, 2011; Murpy & Beggs, 2003; Wang & Berlin, 2010). An earlier study (Oon & Fan, 2017) within an Asian context found that the three commonly accepted constructs of SAS were not highly correlated as have been reported in precursory studies (Murphy & Beggs, 2003; Wang & Berlin, 2010). They recommended that the negatively worded SAS items be deleted or replaced by positively worded items. These recommendations elicited further attention to the formation of SAS constructs of which the present study aims to seek further affirmation on.

Aim of this Study

We explore how Rasch analysis can be used to provide psychometric information to validate Wang and Berlin (2010)'s SAS instrument in China and how such information could help to improve the psychometric quality of this SAS scale.

METHODOLOGY

Survey Instrument

A survey instrument from Wang and Berlin (2010) – *Asian Student Attitudes towards Science Class (ASATSC)* was used in this study to explore the structure of SAS in a Chinese context. It contains 30 items with three constructs: (1) science enjoyment (SE) – the pleasure students feel in science class; (2) science confidence (SC) – the evaluation of students' perceived abilities and capabilities towards science; and (3) importance of science (IS) – the importance of learning activities in science class.

The survey is divided into two sections. The first section sought demographic information on students' genders and school levels. The second section included 16 positively and 14 negatively worded SAS items in a five-point Likert-type response format (1= strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, 5 = strongly agree).

The original ASATSC in English (Wang & Berlin, 2010) was translated by the first and fourth authors into Chinese. Four evaluators were invited to evaluate the translated version via e-mail. The translated survey was sent to the evaluators via email after they agreed to participate in this research. Two postgraduate students specialized in English-Chinese translation proofread the translations, and two academic staff from the University of Macau validated the last version. They evaluated the wording for each item, with an indication of '✓' or '✗' responses and provided suggestions to those they felt inappropriate. The two postgraduate students indicated a 66% of appropriateness. According to their comments, some items were not translated appropriately (e.g., the expression reads colloquial). After careful modification, the last version was re-sent to the two academic staff. They gave positive comments to the revised translations and indicated 95% of agreement on the translations' appropriateness. The final survey used in the current study was produced as a result.

Research ethics for the current study was evaluated and approved by the University of Macau before data collection. All items were divided into three constructs (**Table 1**): (a) the first construct consisting of 13 items was labeled 'Science Enjoyment' (SE), (b) the second construct consisting of eight items was labeled 'Science Confidence' (SC), and (c) the third construct consisting of nine items was labeled 'Importance of Science' (IS).

Table 1. Items to SAS constructs for the current study (Wang & Berlin, 2010)

Construct	Items
Science Enjoyment (SE)	1.I like when the teacher teaches our Science outdoors
	2.In Science class, listening to lectures from the teacher is interesting
	3.In Science class, watching the Science film on TV is boring
	4.My Science class is interesting
	6.I would enjoy school more if there was no Science class
	7.During Science class, I like to read Science posters
	8.I look forward to Science class
	9.In Science class, doing experiments is boring
	10.I do not like Science class
	17.I enjoy reading the Science textbook
	19.I like to do experiment in Science class
	20.I do not like answering the questions in my Science workbook
	28.I do not like field trips in my Science class
Science Confidence (SC)	11.The material in the Science textbook is hard for me
	12.I am afraid to answer the questions in Science class
	15.In Science class, experiments are difficult
	18.I usually understand what is taught in my Science class
	22. Science class is hard for me
	23.It is easy for me to understand the teacher's lectures in Science class
	25.I usually get good scores in Science class
	30. The questions in the Science workbook are easy for me
Importance of Science (IS)	5.In Science class, I learn more Science when I work in a group
	13. Science class provides me with knowledge to use in my daily life
	14.The experiments I do in Science class are useful
	16.In Science e class, Science poster do not help me to learn Science
	21.In my Science class, field trips do not help me to learn Science
	24.The material in the Science textbook help me to learn Science
	26.The questions in the Science workbook do not help me to learn Science
	27.In Science class, watching Science film on TV helps me to learn Science
	29. Science class is a waste of time

Table 2. Demographic information of the student samples (N=1,133)

		Number	**%**
Gender	Male	622	54.90
	Female	511	45.12
School level	7th grade	239	21.11
	8th grade	371	32.71
	9th grade	67	5.91
	10th grade	157	13.91
	11th grade	299	26.42

Participants and Data Collection Procedure

The study was conducted in Guangzhou, China. A total of 230 (53 in Tianhe district, 62 in Panyu district, 80 in Huadu district, and 35 in Yuexiu district), out of the total of 514 secondary schools in Guangzhou (Guangzhou Education Bureau, 2017), were invited to participate in the study. A total of eight secondary schools agreed to participate in this research project. One of them was a non-profit private school and the rest were public schools. Three of them were senior high schools and the rest were junior high schools. Two classes of students from each grade from each participating school completed the survey. A total of 1,133 seventh to eleventh graders who studied science completed the survey (**Table 2**). Twelfth graders were not invited to take this survey because they needed to prepare for the university entrance examination. The participants and the participating schools were assured that the collected responses would be kept confidentially and the consolidated data would be used for research purposes only.

The survey was conducted in June 2016 by the researchers in Guangzhou, China. Letters of invitation were sent to the eight schools after IRB approval was obtained from the University of Macau. Students' surveys were sent to the science teachers of the eight schools that agreed to participate in this study. A face-to-face briefing session was held by the first author to the students before the distribution of the surveys. Each student had 15 minutes to answer

the survey and was confirmed of response confidentiality and that the results would only be used for research purposes. Science teachers collected the completed surveys and returned the completed survey to the researcher.

Analyses

Identifying constructs important to SAS requires evaluating SAS data in a way that assess the invariant relation between student agreeability and SAS item difficulty. It takes the form of:

$$\ln[P_{ni} / (P_{ni-1})] = B_n - D_i$$

which says that the log-odds of observed success for student *n* on item *i* is equal to the difference between the estimate B of student *n*'s ability and the difficulty estimate D of item *i* (Andrich, 2010; Rasch, 1960; Wright & Masters, 1982). SAS data must reasonably fit a model of this kind in order to explore meaningfully constructs valid to the measurement of SAS. Estimates of agreeability measure B and SAS item calibration D are expressed in common units which allows objective measurement to be made on constructs across relevant samples. All measures are estimated with individual uncertainty and model fit statistics. This Rasch measurement model specifies that the degree of precision is supported by the intended decision process (Linacre, 1993).

RESULTS

Model Fit and Data Reliability

Each participant's responses on the 30 items were subjected to Rasch analysis using WINSTEPS software (Version 3.81.0) to measure whether the data fit the Rasch model. In Rasch analysis, fit statistics is reported as Infit/Outfit MNSQ (mean squares) which helps investigators to check how suitable the empirical data meets the requirement of the Rasch model (Bond & Fox, 2001, *p*.41). It provides information on whether item estimates are meaningful to study the latent trait (Bond & Fox, 2001, *p*.27). The content of items does not objectively define the construct if the fit statistics does not stay within the acceptable range (Oon & Subramaniam, 2011, *p*. 127). The acceptable value of Infit/Outfit MNSQ ranging from .60 to 1.40 was regarded as fit to the Rasch model (Romine & Walter, 2014). In the current study the content refers to SAS.

The Infit/Outfit MNSQ is a transformation of residuals which is used to illustrate the difference between the model and empirical data. The Infit statistics (weighted) is sensitive to students' performances close to the items value. The Outfit statistics (unweighted) is *sensitive to the influence of outlying scores* (Bond & Fox, 2001, *p*. 43). **Table 3** shows the fit statistics of the data. All items were reported to have reasonably good fit to the Rasch model (i.e., Infit/Outfit MNSQ values ranging from .60 to 1.40).

Table 3. Rasch statistics for the SAS items

Construct	Item Description	Item	Total Score	Total Count	Measure	Model S.E.	Infit MNSQ	Oufit MNSQ
Science Enjoyment	1. I like when the teacher teaches our Science outside	I1	7860	2022	-.67	.02	1.19	1.18
	2. In Science class, listening to lectures from the teacher is interesting	I2	7190	2018	-.38	.02	.90	.91
	3. In Science class, watching the Science film on TV is boring	I3R	4819	2022	.49	.02	1.20	1.23
	4. My Science class is interesting	I4	7448	2022	-.48	.02	.87	.88
	5. I would enjoy school more if there was no Science class	I6R	4517	2016	.61	.02	1.31	1.34
	6. During Science class, I like to read Science posters	I7	7010	2012	-.32	.02	.86	.88
	7. I look forward to Science class	I8	7072	2014	-.34	.02	.89	.89
	8.In Science class, doing experiments is boring	I9R	4070	2015	.80	.02	1.12	1.20
	9. I do not like Science class	I10R	4366	2012	.67	.02	1.13	1.17
	10. I enjoy reading the Science textbook	I17	7226	2017	-.40	.02	.84	.84
	11. I like to do experiment in Science class	I19	7634	2016	-.57	.02	.98	.98
	12. I do not like answering the questions in my Science workbook	I20R	5568	2015	.21	.02	1.06	1.08
	13. I do not like field trips in my Science class	I28R	4481	1994	.60	.02	1.17	1.17
Science Confidence	14. The material in the Science textbook is hard for me	I11R	5463	2012	.24	.02	1.06	1.07
	15. I am afraid to answer the questions in Science class	I12R	5340	2014	.29	.02	1.07	1.09
	16. In Science class, experiments are difficult	I15R	4792	2010	.49	.02	.92	.96
	17. I usually understand what is taught in my Science class	I18	7263	2016	-.41	.02	.78	.79
	18. Science class is hard for me	I22R	5170	2017	.35	.02	1.09	1.12
	19. It is easy for me to understand the teacher's lectures in Science class	I23	6462	2013	-.11	.02	.82	.82
	20. I usually get good scores in Science class	I25	6394	2001	-.10	.02	.88	.89
	21. The questions in the Science workbook are easy for me	I30	6844	2009	-.26	.02	1.23	1.61
Importance of Science	22.In Science class, I learn more Science when I work in a group	I5	7285	2022	-.41	.02	.90	.91
	23. Science class provides me with knowledge to use in my daily life	I13	7751	2012	-.63	.02	.92	.93
	24. The experiments I do in Science class are useful	I14	7605	2011	-.57	.02	.90	.90
	25. In Science class, Science poster do not help me to learn Science	I16R	4816	2014	.48	.02	.93	.97
	26. In my Science class, field trips do not help me to learn Science	I21R	4684	2018	.54	.02	1.14	1.15
	27. The material in the Science textbook help me to learn Science	I24	7610	2014	-.56	.02	.80	.80
	28. The questions in the Science workbook do not help me to learn Science	I26R	4843	1995	.46	.02	.91	.91
	29. In Science class, watching Science film on TV helps me to learn Science	I27	7751	1992	-.67	.02	.88	.87
	30. Science class is a waste of time	I29R	4374	2004	.66	.02	1.14	1.34
	MEAN		6123.60	2012.30	.00	.02	1.00	1.03
	S.D.		1303.30	7.80	.49	.02	.14	.19

Person and item reliabilities were .55 and 1.00, respectively. A lower person reliability (<.80) indicates the SAS items may not be sensitive enough to distinguish students' agreeability level. This indicates that more SAS items, better at targeting students' SAS, would improve the person reliability. The acceptable value of item reliability is above .90. Higher item reliability indicates that the items estimates are reproducible by another relevant sample with similar ability (Bond & Fox, 2001, *p*.32). In other words, high item reliability indicates sufficient sample for the SAS measures (Linacre, 2009).

Differential Item Functioning (DIF)

If there is anything to learn about SAS from the measurement of the current study, the measure must show itself as an invariant pattern that meaningfully locates student responses relative to item and vice versa in a stable manner. It is called the property of invariance. Differential item functioning (DIF) is "*a prima facie evidence of items that help to monitor the loss of invariance of items estimates across testing occasions*" (Bond & Fox, 2001, *p*. 230). To examine whether items function relevantly stable across different subsamples, the current study examined the invariance of items estimates for male and female students. **Table 4** indicates items estimate for male students (indicated as '1') and female students (indicated as '2'). Male and female students had different degrees of agreement for each item. The DIF Contrast for all items reported values less than .50 (*p* > .50) which indicated that none of the items biased against any genders (Linacre, 2009).

Table 4. Differential item functioning for male and female students for the SAS item

Construct	Person Class	Item	DIF Measure	DIF S.E.	Person Class	DIF Measure	DIF S.E.	DIF Contrast	Rasch-Welch *p*
	1	I1	-.69	.03	2	-.64	.03	-.05	0.28
	1	I2	-.38	.03	2	-.38	.03	.00	1.00
	1	I3R	.49	.03	2	.47	.03	.02	0.59
	1	I4	-.48	.03	2	-.46	.03	-.02	0.58
	1	I6R	.61	.03	2	.61	.03	.00	1.00
	1	I7	-.34	.03	2	-.29	.03	-.06	0.16
Science Enjoyment	1	I8	-.37	.03	2	-.30	.03	-.07	0.10
	1	I9R	.85	.03	2	.75	.03	.10	0.02
	1	I10R	.69	.03	2	.64	.03	.06	0.17
	1	I17	-.45	.03	2	-.34	.03	-.11	0.00
	1	I19	-.57	.03	2	-.55	.03	-.02	0.62
	1	I20R	.24	.03	2	.17	.03	.07	0.05
	1	I28R	.56	.03	2	.65	.03	-.09	0.04
	1	I11R	.31	.03	2	.17	.03	.14	0.00
	1	I12R	.39	.03	2	.17	.03	.22	0.00
	1	I15R	.53	.03	2	.44	.03	.09	0.03
Science Confidence	1	I18	-.46	.03	2	-.36	.03	-.10	0.01
	1	I22R	.37	.03	2	.33	.03	.04	0.24
	1	I23	-.18	.03	2	-.03	.03	-.15	0.00
	1	I25	-.15	.03	2	-.05	.03	-.09	0.01
	1	I30	-.26	.03	2	-.26	.03	.00	1.00
	1	I5	-.46	.03	2	-.36	.03	-.09	0.02
	1	I13	-.60	.03	2	-.67	.03	.07	0.11
	1	I14	-.57	.03	2	-.57	.03	.00	1.00
Importance of Science	1	I16R	.52	.03	2	.44	.03	.08	0.05
	1	I21R	.54	.03	2	.52	.03	.02	0.56
	1	I24	-.56	.03	2	-.56	.03	.00	1.00
	1	I26R	.49	.03	2	.41	.03	.09	0.03
	1	I27	-.71	.03	2	-.61	.03	-.10	0.02
	1	I29R	.66	.03	2	.66	.03	.00	1.00

Note: Person class 1= male students, person class 2= female students

Table 5. Summary of category structure of 5-point rating scales for the student SAS scale

Category	Observed Count (%)	Average Measure	Outfit MNSQ	Threshold Calibration
1	9991(17)	- .42	1.09	NONE
2	11576(19)	- .31	.81	-.50
3	14571(24)	.00	.96	-.36
4	14303(24)	.29	.84	.14
5	9928(16)	.42	1.19	.72

Effectiveness of Response Categories

Rating scale analysis is used to examine the usage of each of the Likert scale category. It helps to examine whether rating categories are appropriately and optimally used.

A criterion of Rasch was used to verify the effectiveness of each of the 5-point response categories. A minimum of ten observations were made and the outfit MNSQ for each category reported values below 2.00 (**Table 5**).

The average measure increased monotonically from -.42 (Strong Disagree) to .42 (Strong Agree). The threshold calibration also increased monotonically (**Table 5**). The results suggested that each category worked optimally as intended (Linacre, 2002). Although each category had an obvious peak (**Figure 1**), the distance of threshold calibration between Category 2 (Disagree) and Category 3 (Undecided) was .14 which should be considered to be collapsed to increase the reliability of the data (Linacre, 2002).
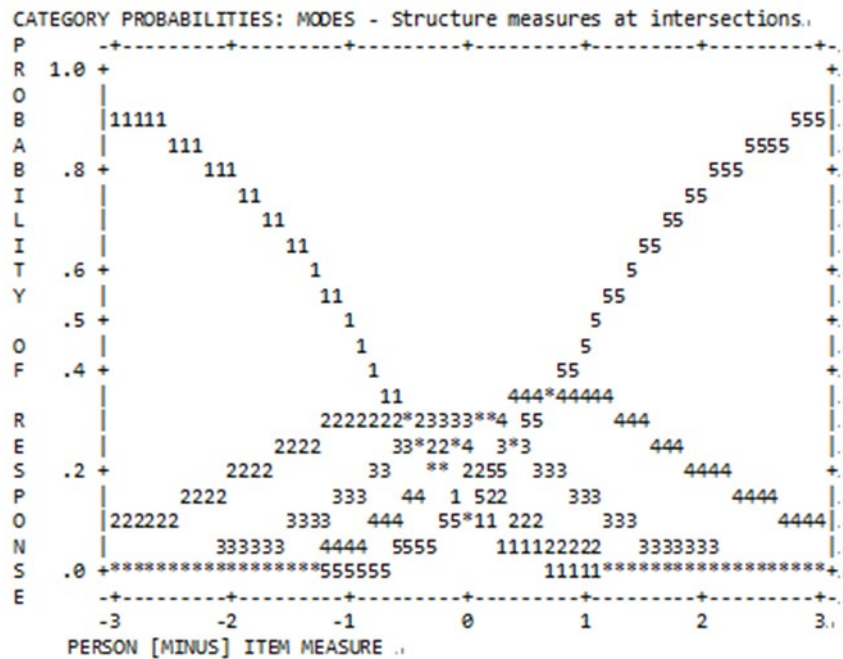
Figure 1. Category probability curves for the 5-piont rating scale

Table 6. Summary of analysis for original and collapsed scales

		Rating scale	
		Original (1-2-3-4-5)	**Revised** (1-2-2-3-4)
Fit statistics			No improvement
Reliabilities	Person	.55	.59
	Item	1.00	1.00
Raw variance explained by measure		30%	28.5%
DIF contrast			No improvement
Linacre's criteria	N>10	✓	✓
	M(q)$_{increase}$	✓	✓
	MS<2	✓	✓
	τ$_{increase}$	×	×
	Curves $_{peak}$	✓	✓

Note. ✓ indicates satisfied; × indicates not satisfied

Improvement on the SAS Measurement

Since the category function did not meet Linacre's criteria requirement, an attempt was made to reorganize the five-point scale (1-2-3-4-5) to four-point scale (1-2-2-3-4). Category 2 and Category 3 were collapsed as one category threshold. **Table 6** summarizes the adequacy of the original and collapsed categories. However, the other Rasch index did not show significant improvement (**Table 6**). The results prompted the use of the original categories for results interpretations.

As afore-mentioned, the low person reliability indicates the need to include SAS items better at targeting the measurement of SAS. We further examined the quality of the items through principal component analysis (PCA) of residuals. Strong evidences indicated that negatively worded items might have confused the students in their responses. The person reliability increased from .50 to .85 after removing all negative items (**Table 7**). Negatively worded items should be treated with caution when measuring the construct of SAS.

Table 7. Summary of analysis for original and only positive worded items

		Original	Only positive worded items
Reliabilities	Person	.55	.85
	Item	1.00	1.00
Dis-attenuated Correlation	SC-IS		0.66
	SC-SE		0.96
	IS-SE		0.98
Raw variance explained by measure		30%	38.2%
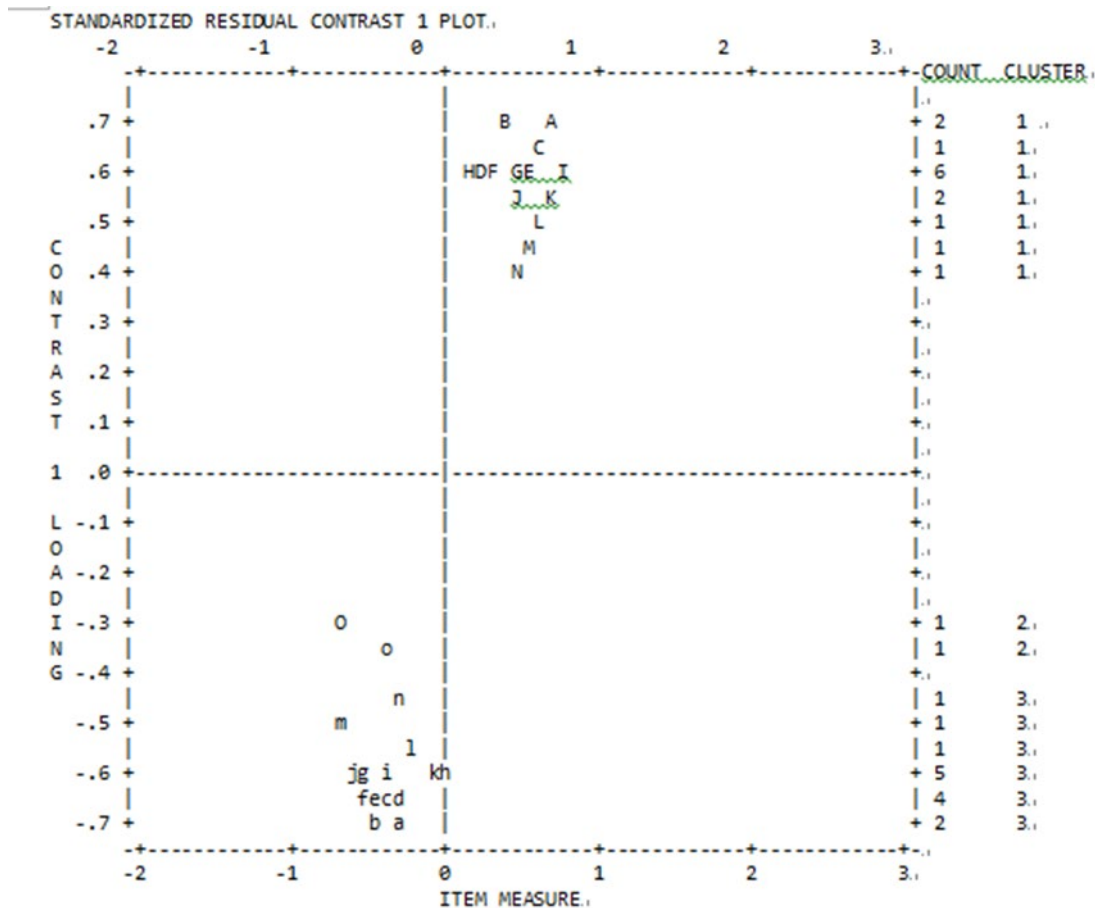
Figure 2. Plot of Residuals loadings for SAS data

For the original items, though all items stayed within the acceptable fit, the variance explained by the Rasch measures was only 30.0% and the first three unexplained variances were 10.0%, 2.0% and 1.6%, respectively in the PCA of residuals. The result indicates the existence of a secondary dimension (noise) in the SAS construct. We further examined which items contributed to the noise through the exploration of the residual loadings plot (**Figure 2**).

Figure 2 shows Item A to Item N, the negatively worded items had factor loadings greater than .40. In contrast, Item a to Item k, the positively worded items, had factor loadings less than -.60.

Table 7 summarizes the analyses between all original items and the positive items. The person reliability increased from .55 to .85 through the removal of all negative items. The variance explained by the Rasch measures then increased from 30.0% to 38.2% (**Table 7**), and the first three unexplained variances decreased to 1.9, 1.7 and 1.5, respectively. SC and SE items as well as IS and SE items showed high correlations of .96 and .98, but the dis-attenuated correlation between SC and IS was .66 (**Table 7**), which revealed these two components were not highly correlated in measuring SAS (Linacre, 2014). This strongly suggested that SC and IS were not consistent in measuring SAS.

Figure 3 shows the person and item estimates as a map. Each "#" denotes 31 students and each "." denotes 1 to 30 students. Students at the top of **Figure 3** liked science more than those at the bottom, items at the top of the figure were more difficult SAS science statements for students to agree with and items towards the bottom were easier to
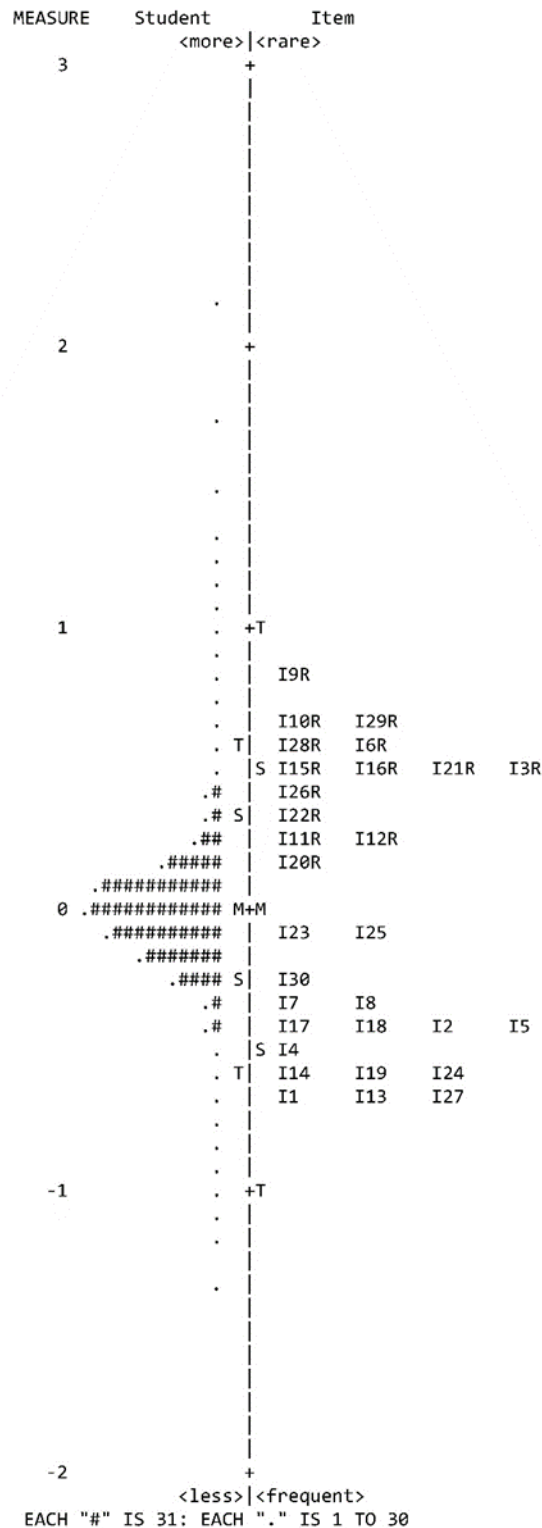
```
MEASURE     Student         Item
               <more>|<rare>
   3                   +
                       |
                       |
                       |
                       |
                       |
                       |
                       |
                       |
                 .     |
   2                   +
                       |
                 .     |
                       |
                 .     |
                 .     |
                 .     |
                 .     |
   1             .     +T
                 .     |
                 .     |   I9R
                 .    .|   I10R    I29R
                 .  T|     I28R    I6R
                 .     |S  I15R    I16R    I21R    I3R
                .#     |   I26R
                .# S|      I22R
               .##     |   I11R    I12R
             .####     |   I20R
          .##########  |
   0   .############ M+M
           .######### |    I23     I25
             .####### |
              .#### S|     I30
               .#     |    I7      I8
               .#     |    I17     I18     I2      I5
                 .     |S  I4
                 .  T|     I14     I19     I24
                 .     |    I1      I13     I27
                 .     |
                 .     |
                 .     |
  -1             .     +T
                 .     |
                 .     |
                       |
                 .     |
                       |
                       |
                       |
                       |
                       |
                       |
  -2                   +
               <less>|<frequent>
       EACH "#" IS 31: EACH "." IS 1 TO 30
```

Figure 3. Item and Person measures estimated on a calibrated linear scale

be agreed with. Students staying closer to the top of **Figure 3** were more in favor of science than those near the bottom. All positive items in **Figure 3** lie beneath the middle whilst all negative items were above, meaning that most of the sampled students found it more difficult to agree with the negative items. This indicated that these negative items were rarely endorsed by the sampled students in the SAS measurement.
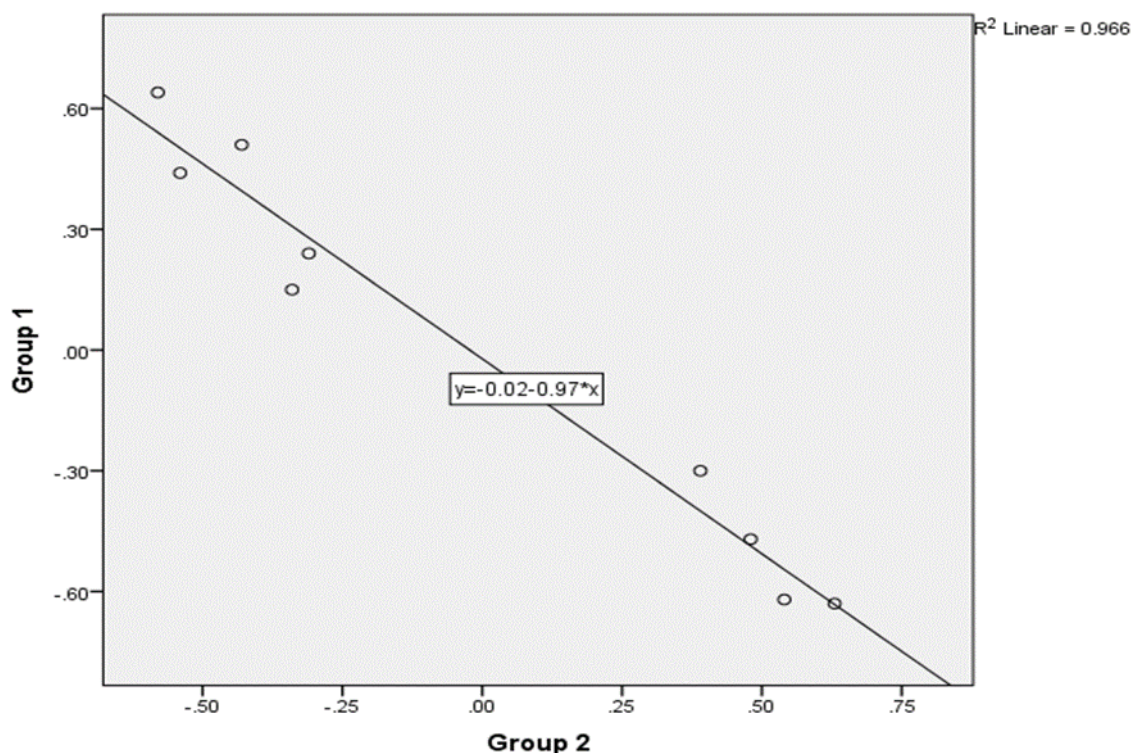
Figure 4. Scatter-plot for contrast loadings between Group 1 and Group 2 randomly split students

To examine whether the contrast factor loadings between SC and IS remains invariant in small sub-samples, five groups were randomly chosen from the 1,133 student sample. We ran not only the PCA of the residuals on these five sub-samples but also plotted simple scatter-plot for contrast loadings between SC and IS. There were ten possible pairwise graphs in total (e.g., Group 1 vs. Group 2, Group 1 vs. Group 3, Group 4 vs. Group 5, etc.). One of the ten representative graphs (See **Figure 4**) showed the contrast loadings between Group 1 and Group 2 and these two groups produced a loading overlapped on 96.6%. The ranges of shared variances across the ten sampled scatter-plots ranged from 74.2% to 97.7%. Although three of the shared variances were less than 80%, due to the small sample size, most of them remained linearly invariant which strongly showed SC and IS constructs that did not correlate well with each other in the SAS measure. This reminds us to be cautious with these three commonly used SAS constructs and suggests that the results to these two constructs should be interpreted separately.

DISCUSSION

Insights from Rasch Analysis for Improving SAS Rating Scales

The strength of Rasch models, as described by Embretson and Reise (2000, pp. 324-325), lies in their support for the study of item-positioning effects, cultural differences in item functioning, assessment of individual-level response consistencies, interpretation of the meaning of scores, and exposure of poorly functioning or poorly conceived items. However, a research into science education literature reveals that Rasch Model has rarely been used in validating SAS rating scales. A few scholars (e.g., Boone, Staver, & Yale, 2014; Liu, 2010) have recently encouraged science education researchers to use Rasch model for developing psychometrically better instruments in science education research. We feel that though Rasch analysis has been receiving more attention in science education, many still are unaware of what Rasch analysis can offer and how information from the analysis can be used in improving the psychometric quality of instruments.

The reliability of data depends not only on the suffix of samples and items but also the optimal usage of rating categories (Boone, Townsend, & Stave, 2011). It is suggested that the underutilised categories should be revised (Wright & Linacre, 1992). However, our attempt to collapse the five-point categories to four did not show significant improvement. As a result, the original categories were kept for data analyses and interpretations.

Only very few studies considered DIF analysis in examining the invariance property of a scale. In fact, this property is important in developing measures. If items were found to function differently across sub-samples, the

measurement results could be very misleading (Bond & Fox, 2001). None of the items from Wang and Berlin's (2010) was flagged in Rasch analysis as being not invariant.

The Rasch's model-fit statistics and PCA analysis of residuals provide details on the unidimensionality of the latent trait being measured, that is, attitudes towards science in this study. The residual-based PCA is capable of identifying secondary dimension (e.g., noises) that may plague the data. In the current study, the Rasch-residual-based PCA identified secondary dimension from the negatively worded items. This means that the data collected from the SAS scale may not only measure attitudes towards science but other 'noises' as well. This information is particularly important if scores are to be summed across all items for a total mean for results interpretation.

Recommendation on SAS Constructs Improvement

The PCA of residuals suggested that the negatively worded items might not measure the same construct as the positively worded items did, as evident in the improved reliability index and variance explained by Rasch measures, and a better fit was achieved by omitting the negatively worded items. The negative items might have confused the students in their responses towards the SAS items (Oon & Fan, 2017). The findings suggested that the negatively worded items should be treated with caution when measuring the constructs of SAS. As a result, only positively worded items were included for correlational analysis between SC, SE and IS. It is interesting to find then science confidence (SC) and importance of science (IS) were not correlated well in measuring SAS as compared to many precursory studies using conventional analysis in validating the psychometric properties of SAS instruments (e.g., exploratory factor analysis) (e.g., Jocz, Zhai, & Tan, 2014; Swarat, Ortony, & Revelle, 2012).

There is no consensus among researchers as to which constructs should be included in the measure of SAS (Li, 2013). A recent study by Wan and Lee (2017) with samples from Hong Kong secondary schools classified SAS constructs into the following three domains: cognition, affect, and behavior in a total of seven dimensions, three were related to the constructs in the present study, they are value of science to society (IS), self-concept in science (SC), and enjoyment of science (SE). The selection of SAS often depends on the subjective judgment of researchers. The current study based on Wang and Berlin (2010) included constructs of science enjoyment (SE), science confidence (SC), and importance of science (IS), as reported in Bathgate, Schunn and Correnli (2014), Bryan, Glynn and Kittleson (2011), Murphy and Beggs (2003), and TIMSS 2011 (Martin & Mullis, 2012).

According to Wang and Berlin (2010), importance of science (IS) is recognized as *the extent to which a student thinks their science class to be an important and worthwhile class* (*p*. 2418). In their study, the items of IS were referred to as teaching materials and teaching strategies commonly used in the elementary school science classroom. Often, teachers attempted to increase students' learning interests through various active teaching methods (e.g., experiment, science film, textbook, and group learning) (Gan, 1994; He, 2014; Li, 2013; Ni, 2012), hoping to attract students' attention (Li, 2013) and improve their attitudes towards learning science (Ni, 2012; Wen, 2015). It is important to note that IS pertains to extrinsic motivation.

Science confidence (SC) is used to measure confidence in one's ability and capabilities to learn science. According to Dhindsa and Chung (2003), SC is recognized as the *extent to which student is confident and successful in doing science* (*p*. 911). Lee, Hayes, Seitz, DiStefano, and O'Connor (2016) examined three competing models of motivational constructs and summarized self-efficacy (confidence) as one of the constructs pertaining to intrinsic motivation.

Science enjoyment (SE) has often been referred to as intrinsic motivation, as '*doing something because it is inherently interesting or enjoyable*' (Ryan & Deci, 2000, *p*. 55 as reported in Palmer, 2005, *p*.1858). That means a person feels pleased when s/he is doing something s/he likes.

Oon and Fan (2017) conducted analyses using three similar motivational constructs as those used in the current study from the existing TIMSS dataset, including *the student like learning science scale, student value science scale, and student confidence in science*. The results showed that *student value science scale* and *student confidence in science* were not correlated well, while *student like learning science scale* and *student confidence in science* were consistent with each other in the measure of SAS. The authors gave a theoretical explanation that both *student confidence in science* and *student like learning science* are intrinsic aspects of motivation, while *student value science* is extrinsic. Thus, they are conceptually different. According to the results mentioned above, the authors suggested to combine *student like learning science scale* and *student confidence in science* into one internal dimension and treat *student value science scale* as an external dimension, this is affirmed in the findings of the current study that IS and SC did not measure SAS coherently, as evident both statistically and theoretically, i.e., one is perceived intrinsically while the other extrinsically. On the other hand, SC and SE can be combined into one as internal factor. In conclusion, IS should be regarded as external factor while SC & SE as internal factors. Results to these two factors should be interpreted separately.

CONCLUSIONS

This study sets to explore how Rasch analysis can be used to provide psychometric information to validate Wang and Berlin (2010)'s SAS instrument in China and how such information could help to improve the psychometric quality of this SAS scale.

The analysis of PCA of residuals suggested that the negatively worded items did not measure the same construct of SAS as those positively worded items did as all negatively and positively framed items clustered separately. This finding corroborates with other studies (e.g., Conrad, Wright, McKnight, McFall, Fontana, & Rosenheck, 2004) where the measure explained by Rasch increased significantly and the noises decreased sharply in the data after removing the negatively worded items. Bainer and Smith (1999) stated that the negatively worded items do not measure the same underlying construct as the positively worded items do and so the two kinds of items in a same calibration often cause highly incompatible situations. They urged

> *"...be careful when introducing reverse coded or negatively worded items into the instrument. Although this practice has been recommended as a means of offsetting response set biases, there are clear indications in a variety of settings that the responses to the negatively worded items do not measure the same underlying construct as the positively worded items. There may be a substantial correlation between the two variables, as there was in this case, but the combination of the positively and negatively worded items in the same calibration often causes the item fit statistics to have an unexpectedly high proportion of misfitting items." (p. 263).*

Therefore, we recommended that the negatively worded items be replaced by positively worded items.

The statistical findings also prompted us to suggest that SC was not correlated well with IS whilst SC and SE were consistent with each other in the measure of SAS. SC and SE, therefore, can be combined as one internal factor whilst IS can be treated as external factor in the interpretation of students' attitudes towards science.

Limitations of the Study

The sample size (n=1,133) is not representative of all populations of Chinese students in China. The results should be viewed within the province of Guangzhou, China.

The inclusion of the three constructs, namely, science enjoyment, science confidence, and importance of science was based on Wang and Berlin (2010). Though we found the importance of science was not measuring SAS coherently as compared to science enjoyment and science confidence, the importance of other constructs remained unknown. Therefore, the findings should be construed as being indicative rather than confirmatory as not all constructs in the measure of attitudes were incorporated in this instrument.

REFERENCES

Andrich, D. (2010). Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika, 75*(2), 292-308. https://doi.org/10.1007/s11336-010-9154-8

Angell, C., Guttersrud, Ø., Henriksen, E. K., & Isnes, A. (2004). Physics: Frightful, but fun. Pupils' and teachers' views of physics and physics teaching. *Science Education, 88*(5), 683-706. https://doi.org/10.1002/sce.10141

Anwer, M., Iqbal, H. M., & Harrison, C. (2012). Students' attitudes towards science: A case of Pakistan. *Pakistan Journal of Social and Clinical Psychology, 9*(2), 3-9.

Bainer, D. L., & Smith, R. M. (1999). Developing a unidimensional instrument to measure the effectiveness of shool-based partnerships. *Journal of Outcome Measurement*, 3(3), 248-265.

Barmby, P., & Defty, N. (2006). Secondary school pupils' perceptions of physics. *Research in Science & Technological Education, 24*(2), 199-215. https://doi.org/10.1080/02635140600811585

Barmby, P., Kind, P. M., & Jones, K. (2008). Examining changing attitudes in secondary school science. *International Journal of Science Education, 30*(8), 1075-1093. https://doi.org/10.1080/09500690701344966

Bathgate, M. E., Schunn, C. D., & Correnti, R. (2014). Children's motivation toward science across contexts, manner of interaction, and topic. *Science Education, 98*(2), 189-215. https://doi.org/10.1002/sce.21095

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410600127

Boone, W. J., Staver, J. S., & Yale, M. S. (2014). *Rasch analysis in human sciences*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-6857-4

Boone, W. J., Townsend, J. S., & Staver, J. (2011). 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. https://doi.org/10.1002/sce.20413

Brophy, J. (1998). Motivating students to learn. Madison, WI: McGraw Hill.

Bryan, R. R., Glynn, S. M., & Kittleson, J. M. (2011). Motivation, achievement, and advanced placement intent of high school students learning science. *Science Education, 95*(6), 1049-1065. https://doi.org/10.1002/sce.20462

Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, *33*(1), 7-26. https://doi.org/10.1080/09500693.2010.518644

Cavas, P. (2011). Factors affecting the motivation of Turkish primary students for science learning. *Science Education International*, *22*(1), 31-42.

Conrad, K. J., Wright, B. D., McKnight, P., McFall, M., Fontana, A., & Rosenheck, R. (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD scale: Revealing limitations of reverse-scored items. *Journal of Applied Measurement, 52*(192), 15-30.

DeBacker, T. K., & Nelson, R. M. (2000). Motivation to learn science: Differences related to gender, class type, and ability. *The Journal of Educational Research*, *93*(4), 245-254. https://doi.org/10.1080/00220670009598713

Dhindsa, H. S., & Chung, G. (2003). Attitudes and achievement of Bruneian science students. *International Journal of Science Education*, *25*(8), 907-922. https://doi.org/10.1080/09500690305025

Embretson, S. E., & Reise, S. P. (2000). *Multivariate Applications Books Series. Item response theory for psychologists.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Fraser, B. J. (1978). Development of a test of science-related attitudes. *Science Education, 62*(4), 509-515. https://doi.org/10.1002/sce.3730620411

Gan, W. (1994). Lead-in of new Physics lessons in junior middle school. *Curriculum, Teaching Material and Method,* (6), 29-30. (in Chinese)

George, R. (2006). A cross-domain analysis of change in students' attitudes toward science and attitudes about the utility of science. *International Journal of Science Education*, *28*(6), 571-589. https://doi.org/10.1080/09500690500338755

Germann, P. J. (1988). Development of the attitudes toward science in school assessment and its use to investigate the relationship between science achievement and attitudes toward science in school. *Journal of Research in Science Teaching, 25*(8), 689–703. https://doi.org/10.1002/tea.3660250807

Greene, B. A., & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology*, *21*(2), 181-192. https://doi.org/10.1006/ceps.1996.0015

Guangzhou Education Bureau (2017). *2016 Guangzhou Education Statistics Handbook.* Retrieved from http://www.gzedu.gov.cn/gzeduxxgk/0803/201710/06063d43ba114b9a94d335c9481972ca/files/fe00fd7712f34bffb3999f7fdd88082d.pdf (in Chinese)

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science*, *23*(8), 899-906. https://doi.org/10.1177/0956797611435530

He, C. (2014). Analysis of Biology leading-in teaching in junior high school. *Education Teaching Forum,* (21), 146-147. (in Chinese)

Jocz, J. A., Zhai, J., & Tan, A. L. (2014). Inquiry learning in the Singaporean context: Factors affecting student interest in school science. *International Journal of Science Education,* *36*(15), 2596–2618. https://doi.org/10.1080/09500693.2014.908327

Kessels, U., Rau, M., & Hannover, B. (2006). What goes well with Physics? Measuring and altering the image of science. *British Journal of Educational Psychology*, *76*, 761–780. https://doi.org/10.1348/000709905X59961

Lee, C. S., Hayes, K. N., Seitz, J., DiStefano, R., & O'Connor, D. (2016). Understanding motivational structures that differentially predict engagement and achievement in middle school science. *International Journal of Science Education, 38*(2), 192-215. https://doi.org/10.1080/09500693.2015.1136452

Lee, O., & Brophy, J. (1996). Motivational patterns observed in sixth-grade science classrooms. *Journal of Research in Science Teaching*, *33*(3), 303-318. https://doi.org/10.1002/(SICI)1098-2736(199603)33:3<303::AID-TEA4>3.0.CO;2-X

Li, X. (2013). The art and wisdom of leading Biology class in high school. *The New Classroom Learning, (10),* 76. (in Chinese)

Li, X., & Chen, L. (2016). An empirical study of senior high school students' attitudes towards science. *Science Popularization, 11*(2), 31-35. (in Chinese)

Li, Y., & Yao, B. (2014). The development of questionnaire of public attitudes toward science. *Science Popularization, 9(2)*, 60-65. (in Chinese)

Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transactions, 7*(1), 283-284.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M. (2009). *WINSTEPS* (Version 3.69.0) [Computer Software]. Beaverton, OR: Winsteps.com. Retrieved from http:// www.winsteps.com

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

Martin, M. O., & Mullis, I. V. S. (Eds). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center. Boston College.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*, 487-503. https://doi.org/10.1146/annurev.psych.56.091103.070258

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, *80*(4), 514-523. https://doi.org/10.1037/0022-0663.80.4.514

Michael, O. M., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science.* TIMSS & PIRLS International Study Center: Lynch School of Education, Boston College.

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology*, *21*(4), 388-422. https://doi.org/10.1006/ceps.1996.0028

Mullis, I. V. S. & Martin, M. O. (Eds.) (2013). *TIMSS 2015 Assessment Frameworks.* Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/frameworks.html

Murphy, C., & Beggs, J. (2003). Children's perceptions of school science. *School Science Review*, *84*, 109-116.

Ni, D. J. (2012). Application of multimedia in Physics group cooperative learning. *Mathmetics, Physics and Chemistry Learning: Junior Middle School Version, (11)*, 78-79. (in Chinese)

Nolen, S. B., & Haladyna, T. M. (1989). Psyching out the science teacher: Students' motivation, perceived teacher goals and study strategies. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, March. https://doi.org/10.1016/0361-476X(90)90011-O

Nolen, S. B., & Haladyna, T. M. (1990). Personal and environmental influences on students' beliefs about effective study strategies. *Contemporary Educational Psychology*, *15*(2), 116-130.

OECD (Organisation for Economic Co-operation and Development). (2007). *PISA 2006. Science competencies for tomorrow's world*. Paris: Author.

Oon, P. T., & Fan, X. (2017). Rasch analysis for psychometric improvement of science attitudes rating scales. *International Journal of Science Education, 39*(6), 683-700. https://doi.org/10.1080/09500693.2017.1299951

Oon, P. T., & Subramaniam, R. (2011). Rasch modelling of a scale that explores the take-up of physics among school students from the perspective of teachers. In *Applications of Rasch measurement in learning environments researc*h (pp. 119-139). SensePublishers, Rotterdam. https://doi.org/10.1007/978-94-6091-493-5_6

Oon, P. T., & Subramaniam, R. (2013). Factors influencing Singapore students' choice of Physics as a tertiary field of study: A Rasch analysis. *International Journal of Science Education, 35*(1), 86-118. https://doi.org/10.1080/09500693.2012.718098

Oon, P. T., & Subramaniam, R. (2018). Comparative study of middle school students' attitudes towards science: Rasch analysis of entire TIMSS 2011 attitudinal data for England, Singapore and the U.S.A. as well as psychometric properties of attitudes scale. *International Journal of Science Education, 40*(3), 268-290. https://doi.org/10.1080/09500693.2017.1413717

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. https://doi.org/10.1080/0950069032000032199

Palmer, D. (2005). A motivational view of constructivist informed teaching. *International Journal of Science Education, 27*(5), 1853-1881. https://doi.org/10.1080/09500690500339654

Pell, T., & Jarvis, T. (2001). Developing attitudes to science scales for use with children of ages from five to eleven years. *International Journal of Science Education, 23*(8), 847-862. https://doi.org/10.1080/09500690010016111

Potvin, P., & Hasni, A. (2014). Interest, motivation and attitudes towards science and technology at K-12 levels: a systematic review of 12 years of educational research. *Studies in Science Education, 50*(1), 85-129. https://doi.org/10.1080/03057267.2014.881626

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Romine, W. L., & Walter, E. M. (2014). *Assessing the efficacy of the mum as a valid measure for understanding of macroevolution for undergraduate non-science majors.* Biological Sciences Faculty Publications. https://doi.org/10.1080/09500693.2014.938376

Ryan, R. M., & Deci, E. L. (2000). When rewards compete with nature: The undermining of intrinsic motivation and self-regulation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 13-54). San Diego, CA, US: Academic Press. https://doi.org/10.1016/B978-012619070-0/50024-6

Sheldrake, R. (2016). Confidence as motivational expressions of interest, utility, and other influences: Exploring under-confidence and over-confidence in science students at secondary school. *International Journal of Educational Research*, *76*, 50-65. https://doi.org/10.1016/j.ijer.2015.12.001

Simpkins, S. D., Price, C. D., & Garcia, K. (2015). Parental support and high school students' motivation in Biology, chemistry, and Physics: Understanding differences among latino and caucasian boys and girls. *Journal of Research in Science Teaching, 52*(10), 1386-1407. https://doi.org/10.1002/tea.21246

Spall, K., Stanisstreet, M., Dickson, D., & Boyes, E. (2004). Development of school students' constructions of biology and physics. *International Journal of Science Education, 26*(7), 787-803. https://doi.org/10.1080/0950069032000097442

Stokking, K. M. (2000). Predicting the choice of physics in secondary education. *International Journal of Science Education, 22*(12), 1261-1283. https://doi.org/10.1080/095006900750036253

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching, 49*(4), 515–537. https://doi.org/10.1002/tea.21010

Trumper, R. Y. (2006). Factors affecting junior high school students' interest in biology. *Science Education International, 17*(1), 31-48. https://doi.org/10.1007/s10956-006-0355-6

Tuan, H. L., Chin, C. C., & Shieh, S. H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education*, *27*(6), 639-654. https://doi.org/10.1080/0950069042000323737

Wan, Z. H., & Lee, J. C. K. (2017). Hong Kong secondary school students' attitudes towards science: a study of structural models and gender differences. *International Journal of Science Education*, *39*(5), 507-527. https://doi.org/10.1080/09500693.2017.1292015

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. https://doi.org/10.1080/09500690903431561

Weinburgh, M. (1995). Gender differences in student attitudes toward science: a meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching, 32*(4), 387-398. https://doi.org/10.1002/tea.3660320407

Wen, L. (2015). Explore new mode of physics experiment teaching to enhance students' self-confidence. *Science & Technology Vision, (34)*, 183. (in Chinese)

Wright B., & Linacre J. (1992). *Combining and splitting categories*. Rasch Measurement transactions, *6*, 233-235.

Wright, B. D. (1996). Comparing rasch measurement and factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(1), 3-24. https://doi.org/10.1080/10705519609540026

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis. Rasch Measurement*. MESA Press, 5835 S. Kimbark Avenue, Chicago, IL 60637.

Zeyer, A., Bölsterli, K., Brovelli, D., & Odermatt, F. (2012). Brain type or sex differences? A structural equation model of the relation between brain type, sex, and motivation to learn science. *International Journal of Science Education*, *34*(5), 779-802. https://doi.org/10.1080/09500693.2011.635165

http://www.ejmste.com