Journal of Physics: Conference Series

Chemistry teaching self-efficacy: A scale development

To cite this article: A S Tyas *et al* 2020 *J. Phys.: Conf. Ser.* **1440** 012005

Chemistry teaching self-efficacy: A scale development

A S Tyas[1], Senam[1], A Wiyarsi[1], and E W Laksono[1]

[1]Chemistry Education, Universitas Negeri Yogyakarta, Sleman, Indonesia

Corresponding author: anggiasetianing.2018@student.uny.ac.id

Abstract. This study was aimed to develop the Chemistry Teaching Self-efficacy (CT-SE) instrument. The method of this study is an applied survey method. The data of pre-service teachers' CT-SE was obtained through a questionnaire. Three dimensions of self-efficacy were examined in this study, Personal Science Teaching Efficacy belief (PSTE), knowledge efficacy, and teaching efficacy. A group expert confirmed the content validity of the CT-SE with 50 items that had been administered to 137 pre-service teachers. The data were analyzed using the Rasch model that the met criteria of the unidimensionality, item fit, difficulty/ability estimation, reliability, and information function. The results of the Rasch model showed that there are 39 items CT-SE that fit with the Rash model and high reliability for the person's reliability and excellent reliability for the item reliability. The chemistry teaching self-efficacy is acceptable as a good instrument to collect the data. This study suggests that chemistry teaching self-efficacy will prove to be a useful instrument for the measure of pre-service teacher self-efficacy for acid-base topics in chemistry learning.
Keywords: *chemistry teaching self-efficacy instrument, knowledge efficacy, personal science teaching efficacy belief, teaching efficacy*

1. Introduction

Being a chemistry teacher is not merely teaching but also educating. As a teacher, educators not only play a role in conveying knowledge, but are also obliged to conduct evaluations, manage classes, and develop learning tools. Chemistry teacher education has many goals, but from the perspective of pre-service teachers, confidence in his ability to teach chemistry well in class is a real problem. Therefore, it is important to develop chemistry teachers' self-efficacy instruments, which can then be used to measure changes in perceptions of personal chemistry teaching. Teachers' self-efficacy is a teachers' belief in his ability to complete tasks, even though the level of difficulty of the task is high. Self-efficacy is needed by the teacher to make decisions in achieving learning objectives, managing student behavior or holding control in the learning process [1]. The perseverance of a teacher in completing his tasks, such as mastery of learning techniques, mastery of the material, asking students to correct their work. Teacher behavior can increase student involvement in working independently and closeness between teacher and student. Male teachers have higher self-efficacy than female teachers [2], [3].

Self-efficacy is a decision of self-ability related to ability and action [4]. The same meaning is also conveyed by [5] and [6] that self-efficacy refers to the ability to successfully master something. The argues by [7] that self-efficacy can maintain abilities as skills that can be relied upon. Based on this study, self-efficacy refers to one's ability to understand learning so that learning goals are achieved.

Self-efficacy is related to one's self-confidence in academic mastery. As a result of research by [4] revealed that if a person is convinced to believe in himself, he is more persistent in facing difficulties. Self-efficacy acts as a motivator, regardless of whether the achievement should fall substantially, moderately, or minimally from the target set or even exceeding it. It also states that self-efficacy is felt for achieving goals and setting self goals [7]. Hence, it can be concluded that self-efficacy is self-confidence which refers to aspects of self-ability, self-regulation, and the achievement of goals. Self-efficacy is more specific, for example, chemistry students might have high self-efficacy to study acid-base material but not to hydrocarbon material. Self-efficacy is more specific, for example, chemistry students might have high self-efficacy to study acid-base material but not to hydrocarbon material. Self-efficacy beliefs determine how a person feels, thinks, motivates himself, and behaves.

The earliest instrument designed to measure teachers' self-efficacy was developed by [1]. The instrument contained two scales. They are personal teaching efficacy and teaching efficacy. Building on [1] approach, instrument development [2] designed to measure elementary teacher's science teaching efficacy beliefs, the Science Teaching Efficacy Belief Scale (STEBI-A). The STEBI-A contained two scales, Personal Science Teaching Efficacy (PSTE) and Science Teaching Outcome Expectancy (STOE). Developed the STEBI-B, which modified the STEBI-A to measure science teaching efficacy beliefs in pre-service teachers. And then [3] it was re-examined the factor analysis structure supporting the original two scales represented in the STEBI-B, personal science teaching self-efficacy and outcome expectancy. Revised the original SETAKIST for use with K-12 teachers by rewording three items, the scale is named the SETAKIST-R [8]. The elaboration and adaptation from STEBI-B by [9] reveal the confidence of pre-service chemistry teachers to their ability in teaching chemistry at vocational schools. This article aims to develop the Chemistry Teaching Self-efficacy (CT-SE) instrument.

2. Research methods

This study used a survey (instrument development). The sample involved was undergraduate and post-graduate students.

2.1. Development of the chemistry teaching self-efficacy

Instrument development was developed based on the elaboration and adaptation from the personal teaching efficacy and teaching efficacy developed by [1], the Science Teaching Efficacy Belief Scale (STEBI-A) by [2]. The STEBI-B development by [2], which modified the STEBI-A to measure science teaching efficacy beliefs in pre-service teachers, the SETAKIST-R by [8], and the personal science teaching efficacy belief and science teaching outcome expectancy by [9]. To confirm the validity of the content, the researchers outlined a framework theory about the teachers' self-efficacy as a basis for building items and then ascertaining it to the experts. The statements of the instrument are developed appropriately for teaching in vocational context and then translated in Indonesian. There are three dimensions of self-efficacy that were examined in this study, Personal Science Teaching Efficacy belief (PSTE), knowledge efficacy, and teaching efficacy. The response scale for the Chemistry Teaching Self-efficacy ranges from "Strongly Belief = 4" to "Strongly Unbelief = 1". Besides ensuring content validity, a focus group discussion of experts (two chemistry educators) was used to gain face validity. Specifically, experts review all of the items for brevity, clarity, and completeness and come to some level of agreement as to which items should be included in the final scale [10]. Empirical validity was done by giving questionnaires to 137 pre-service teachers including 39 post-graduate students and 98 undergraduate students.

2.2. Data analysis

This study intends to validate the instrument and measure its reliability using the Rasch model by Winstep program. The analysis of instrument characteristics was assessed by the unidimensionality, item fit, difficulty/ability estimation, reliability, and information function. The unidimensionality is the construct validity analysis analyzed by the confirmatory factor analysis (CFA). In confirming whether the data is appropriate for factor analysis or not, the Kaiser-Meyer-Olkin Measure of sampling Adequacy (KMO-MSA), Bartlett test of Sphericity, and the anti-image correlation on varimax (orthogonal) rotation were conducted.

3. Results and Discussion

The development of Chemistry Teaching Self-efficacy was used in this study. There are three dimensions of Chemistry Teaching Self-efficacy, Personal Science Teaching Efficacy belief (PSTE), knowledge efficacy, and teaching efficacy. The amount of the items which were developed are 50 items. The experts have reviewed the content and face validity of the chemistry teaching self-efficacy. For all sentences in the indicator used 'verbs' and need to pay attention to several statement items for each domain. There was no item of chemistry teaching self-efficacy should be added or reduced from this process. The instrument is using the polytomous scales, hence the item characteristics analysis was suitable using the Rasch model. The basic requirement of a Rasch model states the unidimensionality, item fit, difficulty/ability estimation, reliability, and information function.

3.1. Unidimensionality

The unidimensionality aims to test whether each item of the instrument is to measure one variable or one ability only [11]. The unidimensionality was also well-known as the construct validity of the instrument. If the unidimensionality was fulfilled, that the construct validity also fulfilled [9]. The factor analysis was used to obtain a unidimensionality of the instrument. The objective of factor analysis is to identify the relationship among variables by seeking the computational results on Eigenvalue in the matrix of intercultural variance-covariance.

The unidimensionality was initiate by the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO-MSA) and Bartlett's test of Sphericity used in this study, to know the data obtained were appropriate for factor analysis or not. For a measure of sampling adequacy or whether data could be factor well, Pallant [12] suggested that if the KMO-MSA is greater than 0.6 and Bartlett's test of Sphericity must be significant at $\alpha < .05$ the factorability of the correlation matrix is assumed.

As shown in Table 1, the result of the KMO value found to be .711 and it proves that the sample used is adequate (.711 > .05). While the Barlett Sphericity test shows that the variable among this study is correlated (.00 < .05). The data obtained in this study is appropriate for factor analysis on unidimensionality or construct validity.

Table 1. The result of KMO-MSA and barlett sphericity test.

Test	**Chemistry Teaching Self-efficacy**	**Conclusion for Factor Analysis**
KMO-MSA test	.711	Appropriate
The significance value of Barlett Sphericity test*	.00	Appropriate

*statistical significance level of .05

The construct validity aims to determine if the item of the instrument are valid or not according to the empirical data. The construct validity was conducted by the interpretation of the anti-image value on the result of factor analysis. The anti-image correlation value obtained after the KMO-MSA and Barlett test of Sphericity was fulfilled. The factor analysis in proving a construct validity with the anti-image correlation for all items must greater than 0.5, the acceptable level. The result of the anti-image correlation has a value greater than 0.5 for each of the 50 items. Thus, the value of these items has a high contribution to the factor structure of the instrument.

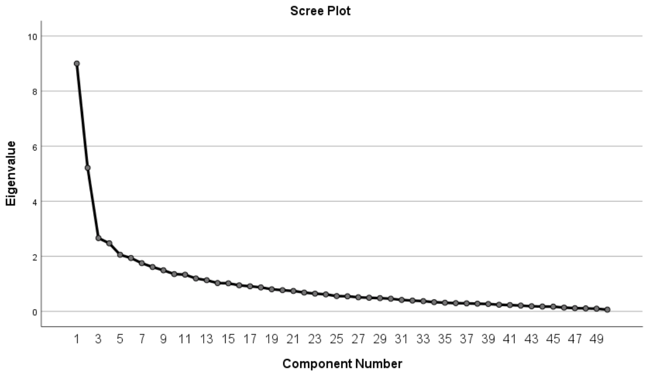

Figure 1. Scree plot of CFA in unidimensionality.

Another way to finding unidimensionality is by the Scree plot. The scree plot is used to describe the illustration of the eigenvalues by the number of components preserve the factors. The unidimensionality could be considered fulfilled if the instrument has a dominant component that measures the ability being tested [13]. The assumption by [14] is that if there is a dominant factor with the cumulative percentage greater than 20%, the unidimensionality can be considered fulfilled. As seen in Figure 1, the scree plot confirms that the instrument of CT-SE consists of 8 factors, which combines the two factors into one. Hence, at least as many as 8 factors were formed with the dominant factor. The percentage of a dominant factor was 23.197%.

The local independence assumption which was carried out and aimed to prove that the participant response toward one item is not affected their response toward the other items of the instrument. The local independence is based on the results of the output person measure sorted from highest to lowest, then processed by making a variance-covariance matrix [15]. The local independence assumption declared fulfilled if when the value under the diagonal line on the variance-covariance matrix is .00. That value indicates the participants' skill in answering the item is not affecting their answering skills toward the other item of the instrument. Table 2, shows that the covariance values on participants' chemistry teaching self-efficacy are approaching .00. The results of this study confirms the idea proposed by [16] that the covariance value is approaching .00, consequently the local independence is fulfilled.

Table 2. Covariance matrix of chemistry teaching self-efficacy.

Columns	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	.062708									
C2	.022462	.00944								
C3	.014638	.005855	.00476							
C4	.0164	.006796	.004805	.005684						
C5	.014038	.004831	.003372	.00346	.003444					
C6	.007708	.003053	.001928	.002247	.001587	.001321				
C7	.008815	.003458	.002428	.002499	.002183	.001001	.001576			
C8	.008162	.003191	.002948	.002672	.001982	.001118	.001449	.002083		
C9	.011031	.004258	.003151	.003122	.002453	.001601	.001622	.001972	.002592	
C10	.034452	.015972	.008278	.01312	.007286	.00456	.005592	.00353	.004128	.044336

3.2. Item fit

In Rasch's measurement, the concept of fit is a quality machine that indicates that the quality of the instruments arranged is adequate. It is also used to assess the meaning of unidimensional constructs, meaning that the fit index helps researchers ensure that Rasch's requirements for dimensions apply empirically. According to [17] the value of the criteria used to check the suitability of the items is as follows, (a) MNSQ Outfit Value received: 0.5 < MNSQ < 1.5; (b) Outfit ZSTD value received: -2.0 < ZSTD < +2.0; (c) Point Measure Correlation (Pt Mean Corr) value: 0.4 < Pt Measure Corr < 0.85.

Table 3. Item fit of the measurement instrument.

Item	Outfit MNSQ	Outfit ZSTD	Pt-Measure Correlation	Conclusion
42	1.80	5.4	0.02	Not fit model
14	1.73	4.7	0.33	Not fit model
12	1.59	3.9	0.45	Not fit model
28	1.58	4.5	0.26	Not fit model
29	1.38	3.3	0.26	Not fit model
41	1.38	2.8	-0.08	Not fit model
33	1.37	2.8	0.07	Not fit model
45	1.32	2.4	0.41	Not fit model
43	1.30	2.3	0.38	Not fit model
47	1.29	2.2	0.29	Not fit model
31	1.27	2.3	0.16	Not fit model
8	1.27	2.0	0.33	Fit model
32	1.23	1.9	0.10	Fit model
4	1.16	1.3	0.37	Fit model
39	1.14	1.2	0.32	Fit model
5	1.12	0.9	0.53	Fit model
48	1.13	1.2	0.13	Fit model
30	1.10	0.9	0.18	Fit model
11	1.09	0.7	0.36	Fit model
10	1.07	0.6	0.39	Fit model
50	1.07	0.7	0.19	Fit model
34	1.01	0.1	0.29	Fit model
25	0.99	0	0.32	Fit model
7	0.99	0	0.35	Fit model
36	0.99	0	0.41	Fit model
37	0.98	-0.1	0.27	Fit model
49	0.95	-0.4	0.26	Fit model
46	0.95	-0.3	0.51	Fit model
44	0.95	-0.4	0.52	Fit model
13	0.94	-0.4	0.56	Fit model
38	0.93	-0.5	0.34	Fit model
20	0.91	-0.7	0.40	Fit model
24	0.90	-0.7	0.39	Fit model
9	0.87	-1.1	0.50	Fit model
19	0.84	-1.3	0.44	Fit model
15	0.80	-1.7	0.61	Fit model
1	0.78	-2.0	0.52	Fit model
35	0.78	-1.9	0.50	Fit model
16	0.75	-2.2	0.58	Fit model
27	0.72	-2.4	0.41	Fit model
2	0.72	-2.6	0.55	Fit model
40	0.71	-2.5	0.42	Fit model
3	0.70	-2.6	0.54	Fit model
6	0.70	-2.6	0.45	Fit Model
17	0.69	-2.8	0.62	Fit model
22	0.69	-2.7	0.50	Fit model
18	0.66	-3.1	0.59	Fit model
21	0.65	-3.1	0.65	Fit model
26	0.63	-3.3	0.58	Fit model
23	0,54	-4.3	0.53	Fit model

According to Table 3, all items are accessed in the results of the item fit order output analysis. There were 11 items that not fit, namely items 12, 14, 28, 29, 31, 33, 41, 42, 43, 45, dan 47. The item fit analysis was used to determine that the items have functioned normally in measuring or not. The received by [18] suggests that the item is statistical fit with the model hence it is considered as a valid item. An item considered fit with the model when at least two criteria of item fit are accepted. The item fit was analyzed using the Winstep program. The result of the analysis showed that 39 items are fit with the PCM-1PL model and can be used for the analysis of measuring pre-service' chemistry teaching self-efficacy.

3.3. Difficulty/ability estimation

The index difficulty provides to find out the correct answer opportunity of a problem at certain ability levels. The parameter of the item difficulty is expressed in logit units. A good instrument item has a range of item difficulty between -2.0 logit and +2.0 logit [16]. An item is considered as a too difficult item if they have index difficulty above +2.00 logit while if they have index difficulty under -2.0 logit it is considered as a too easy item. This study refers on the interpretation of difficulty value following [19] who states that an item categorized very difficult if the value of b (item measure) > 1; difficult .5 ≤ b < 1; medium -.5 ≤ b < .5; easy -.5 ≤ b < -1; and very easy b ≤ -1. Hence based on Table 4, the results of index difficulties of items are well distributed on the very easy, easy, medium, difficult, and very difficult category with the range of the index difficulty between 2.91 and -1.99. According to these range values, it can be said that the item of the instrument has a good difficulty index [19].

Table 4. Item difficulty of chemistry teaching self-efficacy.

Item Number	Difficulty index	Category	Item Number	Difficulty index	Category
11	2.91	Very difficult	38	-.05	Medium
48	2.25	Very difficult	22	-.14	Medium
13	2.04	Very difficult	41	-.22	Medium
27	1.87	Very difficult	40	-.29	Medium
24	1.59	Very difficult	4	-.36	Medium
15	1.41	Very difficult	26	-.36	Medium
23	1.41	Very difficult	37	-.38	Medium
46*	1.41	Very difficult	42*	-.43	Medium
36	1.15	Very difficult	21	-.45	Medium
49	.99	Difficult	7	-.50	Medium
5	.66	Difficult	12*	-.60	Easy
14*	.60	Difficult	29*	-.67	Easy
32	.58	Difficult	50	-.67	Easy
45*	.42	Medium	47*	-.85	Easy
2	.40	Medium	20	-.92	Easy
33*	.40	Medium	1	-1.02	Very easy
39	.31	Medium	18	-1.02	Very easy
30	.29	Medium	35	-1.17	Very easy
44	.18	Medium	10	-1.30	Very easy
3	.16	Medium	43*	-1.30	Very easy
34	.16	Medium	9	-1.32	Very easy
6	.13	Medium	17	-1.48	Very easy
8	.02	Medium	28*	-1.91	Very easy
19	.02	Medium	25	-1.93	Very easy
31*	-.03	Medium	16	-1.99	Very easy

*not fit model

3.4. Reliability

The Rasch measurement model provides indices that help the investigator to determine whether there are enough items spread along the continuum, as opposed to clumps of them, and enough spread of ability among persons to make those decisions. The person reliability index indicates the replicability of person ordering we could expect if this sample of persons were given another a parallel set of items measuring the same construct. The reliability index is between 0 and 1, where 0.8 or above is strongly acceptable. In further determining the reliability of the instrument, person separation and item separation are evaluated. Person separation is an estimate of the extent to which one can differentiate persons on the measured variable, whereas item separation estimates the extent to which all participants are able to answer all levels of difficulty of items. As a rule of thumb, [20] referred to a separation index of higher than 2 as satisfactory separation. Lower values of separation indicate redundancy in items and less variability of persons on the trait.

Table 5. Summary statistics of person and item reliability.

	Infit		Outfit				
Parameter (N)	**MNSQ**	**ZSTD**	**MNSQ**	**ZSTD**	**Separation**	**Reliability**	**Category**
Persons (137)	1.02	-0.2	1.02	-0.2	2.45	0.86	High
Items (50)	1.00	-0.1	1.02	0.0	6.81	0.98	Excellent

The instrument reliability was analyzed according to the person and item analysis. The person's reliability value is found 0.86, indicating high reliability with a 2.45 separation index while the item reliability value is 0.98, indicating excellent reliability with a 6.81 separation index (Table 5). The value of person reliability indicates that there is 86% consistency of the participants' responses toward all of the items in the instrument. While the value of item reliability indicates that there is 86% that there is 98% certainty of the consistency items in obtaining the same result repeatedly.

3.5. Information function

The information function describes how well each level of capability is estimated [21]. Information function was used to further description of the reliability coefficient of the overall test of items. Figure 2 shows that the maximum IF value of chemistry teaching self-efficacy from an instrument with 50 items is found to be 24.375. The information function graph shows that large information can be obtained at a measured value between -2 and 0; this indicates the statement used is not too difficult (small logit value) and can provide good information for individuals with a slightly lower ability than moderate abilities.

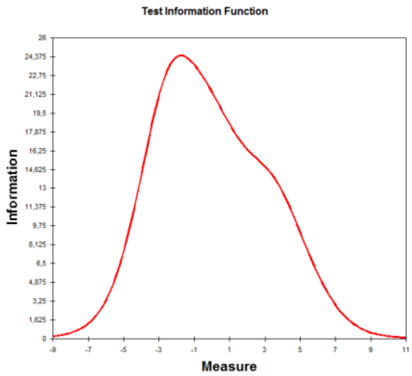

Figure 2. Test information for CT-SE.

The initial CT-SE consists of 50 items that contain the three domains. The final version of CT-SE has 39 items that fit with the Rash model. The number of items on each chemistry teaching self-efficacy was well distributed on 13 items of Personal Science Teaching Efficacy belief (PSTE) dimension, 11 items on knowledge efficacy dimension, and 15 items covering on teaching efficacy dimension. The reliability of person and item are found in high and excellent category respectively. The item difficulty was well distributed on a very easy, easy, medium, difficult, and very difficult category. The IF test indicated that the statement used is not too difficult (small logit value) and can provide good information for individuals with a slightly lower ability than moderate abilities.

4. Conclusion

The result of this study showed that Chemistry Teaching Self-efficacy (CT-SE) has high content and construct validity. This suggests that CT-SE is the potential to be a useful instrument for chemistry teachers and researchers for measuring the pre-service teacher chemistry teaching self-efficacy in an acid-base topic.

References

[1] Gibson S and Dembo M H 1984 *J. Educ. Psy.* **76** 569-82 https://doi.org/10.1037/0022-0663.76.4.569
[2] Riggs I M and Enochs L G 1990 *Sci. Educ.* **74** 625-37 https://doi.org/10.1002/sce.3730740605
[3] Bleicher R E 2004 *Sch. Sci. Math.* **104** 383-91 https://doi.org/10.1111/j.1949-8594.2004.tb18004.x
[4] Bandura A 2012 *J. O. M.* **38** 9-44 http://doi.org/10.1177/0149206311410606
[5] Caprara G V, Vecchione M, Alessandri G, Gerbino M and Barbaranelli C 2011 *Bri. J. Educ.Psy.* **81** 78-96 http://doi.org/10.1348/2044-8279.002004
[6] Judge T A, Jackson C L, Shaw J C, Scoot B A and Rich B L 2007 *J. App. Psy.* **92** 107-27 http://doi.org/10.1037/0021-9010.92.1.107
[7] Wood R and Bandura A 1989 *J. Pers. Soc. Psy.* **56** 407-15 http://doi.org/10.1037/0022-3514.56.3.407
[8] Pruski L A, Riggs R A, Fordtran C W, Cornell J E, Blanco S L, Barbola J M and Lichtenstein M J 2013 *J. Sci. Teac. Educ.* **24** 1133-56 http://doi.org/10.1007/s10972-013-9351-2
[9] Wiyarsi A 2018 *Indo. J. Sci. Educ.* **6** 14-23 http://doi.org/10.26714/jps.6.1.2018.14-23
[10] Sangoseni O, Hellman M and Hill C 2013 *Inter. J. All. Heal. Sci. Prac.* **11** 1-12
[11] Reckase M D 1979 *J. Educ. Stat.* **4** 207–30 http://doi.org/10.3102/10769986004003207
[12] Pallant J F 2000 *J. Pers. Ass.* **75** 308–37 http://doi.org/10.1207/s15327752jpa7502_10
[13] Guler N, Uyanik G K and Teker G T 2013 *Inter. Assoc. Soc. Sci. Res.* **2** 1-6
[14] Brown R 2016 *J. Lu. Pul. Resp. Res.* **3** http://doi.org/10.15406/jlprr.2016.03.00076
[15] Greiff S, Wüstenberg S, Molnár G, Fischer A, Funke J and Csapó B 2013 *J. Educ. Psy.* **105** 181-15 http://doi.org/10.1037/a0031856
[16] Hambleton R K and Swaminathan H 1991 *Fundamental of Items Response Theory* (Newbury Park: Sage Publications) pp 10-12
[17] Boone W J, Staver J R and Yale M S 2014 *Rasch Analysis in The Human* Sciences (London: Springer) Chapter 8 pp 164-175
[18] El-Korashy A F 1995 *Educ. Psy. Mea.* **55** 753-63 http://doi.org/10.1177%2F0013164495055005006
[19] Adedoyin O O and Mokobi T 2013 *Inter. J. Asi. Soc. Sci.* **3** 992-1011
[20] Fisher W P 2007 *Ras. Mea. Trans.* **21** 1095
[21] Baker B B and Kim S 2017 *The Basic of Item Response Theory Using R* (London: Springer) Chapter 6 pp 105

Acknowledgment

Thanks to the post-graduate of Yogyakarta State University that have conducted collaborative research between countries and to the reviewers for the suggestions and feedback.