IRJE |Indonesian Research Journal in Education|
|Vol. 7| No. 1|June|Year 2023|

Validating the Academic Writing Creativity and Self-Efficacy Scale: A Rasch Model Analysis

HERRI MULYONO[1*], SRI KUSUMA NINGSIH[2], AND BEAUTY SHOLEHA RAUFI[3]

Abstract

The current study aimed to examine the psychometric characteristics of an instrument used to evaluate academic writing creativity and self-efficacy scale among pre-service teachers, instructors, and school teachers. Using a convenience sampling, an adapted and expanded version of Mitchell's (2021) situated academic writing self-efficacy (SAWSES) questionnaire was distributed to 655 participants with three different occupation statuses, such as pre-service teacher (N = 315, 485%), TVET instructor (N = 80, 12%), and school teacher (N = 260, 40%). Unfortunately, due to the presence of outliers, of 655 data, only 582 data were used for Rasch statistical analyses, including the analysis of (a) unidimensionality, (b) reliability for the item and the person separation, (c) rating scale, (d) item bias, and (e) item differential function. The analysis of Differential Item Functioning (DIF) resulted that the academic writing creativity and self-efficacy scale could potentially be biased when employed to collect the data from participants with differing teaching statuses and educational backgrounds. These findings thus suggest that caution should be taken when using the Academic Writing Creativity and Self-Efficacy Scale, as highlighted by the results of the current study.

Keywords
academic writing, Rasch, self-efficacy, writing, writing creativity

Article History
Received 16 April 2023
Accepted 3 June 2023

How to Cite
Mulyono, H., Ningsih, S.K., & Raufi, B. S (2023). Validating the academic writing creativity and self-efficacy scale: A rasch model analysis. *Indonesian Research Journal in Education |IRJE|*, *7*(1), 121 – 135. https://doi.org/10.22437/irje.v7i1.24789

[1,] University of Muhammadiyah Prof. DR. HAMKA, Indonesia, Corresponding author: hmulyono@uhamka.ac.id
[2, 3] University of Muhammadiyah Prof. DR. HAMKA, Indonesia

Introduction

Generally, academic writing is viewed as a convention of writing style in an academic society that it involves a specific written communication style in delivering the writer's ideas, opinions, or arguments in a clear, concise, and comprehensible writing structure (Joshi et al., 2022). According to McCarthy (1987), writer's academic writing ability reflects their ability in developing the writing ideas and delivered them in coherent and cohesion explicitly connected sentences. More importantly, scholars who have a sufficient academic writing competences would be able to use texts and language to discuss knowledge (Hyland, 2013) or academic-related topics throughout their critical perspective (Wingate & Tribble, 2012) and rationale (Odena & Burgess, 2015).

Many authors assert that academic writing not only involves criticality and rationality of the writers, but also creativity (Badley, 2019; Dobele & Veer, 2019; Murray, 2013; Odena & Burgess, 2015). In other words, a good academic writers should be able to collect expressions and use words creativity to present their ideas (Groot et al., 2019). In this context, Allison (2004) perceives creativity in academic writing as the concept of proposing insightful and reliable problem-solving as well as creating a connection between the knowledge and inquiry field (Allison, 2004). In a similar vein, Cremin (2017), suggests that academic writing creativity is a reflection of the writer's opportunity to express their ideas by engaging in their self-developed process of writing and reflecting on their life experiences while being able to share their knowledge. He also emphasizes that creative academic writing played a critical way and could be used to reflect ones' level of writing confidence and self-efficacy. In this context, academic writing self-efficacy reflects the belief and confidence in the writers' writing capabilities that enable them to produce a piece academic text that complies with the writing convention in scientific community (Kiriakos & Tienari, 2018). The academic writing self-efficacy is perceived as the phase in which the writer could see themselves as a competent and experienced writer, therefore, lots of attempts were performed to develop the finest version of academic writing (Van de Poel & Gasiorek, 2012).

Many scholars have also conducted a study to explore the variable of academic writing self-efficacy with other variables related (Alberth, 2019; Callinan et al., 2018; Chen & Zhang, 2019; Collado et al., 2023; Huerta et al., 2017; Meza & González, 2020; Mickwitz & Suojala, 2020; Mitchell et al., 2021; Mitchell & McMillan, 2018; Pajares, 2003; Plakhotnik & Rocco, 2016; Ruegg, 2018; Teng et al., 2018; Wei et al., 2022; Zumbrunn et al., 2020). For instance, a study by Van de Poel and Gasiorek (2012), investigated the effect of self-efficacy on the academic writing performance of fifty-three EFL university students through an academic writing course for over two years in Dutch. Their study found that after receiving the course program, students could significantly see themselves more confident with their understanding as competent and experienced writers following the study by Pajares (2003), that self-efficacy towards academic activities could be beneficial and detrimental influence depends on its perceived.

However, in our knowledge, there is little evidence concerning the instruments used to examined academic writing creativity and self-efficacy, except one developed by Mitchell

(2021) to measure academic writing self-efficacy in particular setting. This current study adapted and expanded upon Mitchell's (2021) scales to assess the creativity and self-efficacy of Indonesian instructors and teachers when engaged in academic writing activities. The primary research question addressed in this study was: Is the academic writing creativity and self-efficacy scale reliable to use when evaluating pre-service teachers', TVET instructors', and schoolteachers' creativity and self-efficacy in academic writing activities?

Methodology

Research design, site, and participants

A total of 655 participants with three different occupation statuses, such as pre-service teacher (N = 315, 485%), TVET instructor (N = 80, 12%), and school teacher (N = 260, 40%) completed the study survey. There were 465 females (71%) and 190 males (29%). Many of the participants are university students of teacher training and education faculty (N = 315, 48%), or graduated from the university with a diploma degree (N = 8, 1%), bachelor degree (N = 290, 44%), master degree (N = 40, 6%), and doctoral degree (N = 2, 1%).

Table 1. *Tabulation codes*

Demography	Description	Code	N = 655	N = 582
Status	Pre-Service Teacher (H)	H	315	274
	TVET Instructor (I)	I	80	80
	School Teacher (J)	J	260	228
Gender	Female (F)	F	465	415
	Male(M)	M	190	167
Background of Education	Student (V)	V	315	274
	D3 (W)	W	8	8
	S1 (X)	X	290	260
	S2 (Y)	Y	40	38
	S3 (Z)	Z	2	2

Instrument

The current study adopted and expanded the questionnaire offered by Mitchell (2021), aimed to examine the level of teachers' creativity and self-efficacy in academic writing in Indonesia. Mitchell offered an academic writing self-efficacy scale called situated academic writing self-efficacy (SAWSES) grounded by Bandura's self-efficacy theory and a model of constructed writing (see Bandura, 1978, 1994; Bandura et al., 1997; Bandura & Adams, 1977). The thirty-four questionnaire items were developed using a five-point Likert scale divided into four sections, such as the writing essential scale, rational reflective writing scale, creativity identity scale, and technology use to support writing and writing creativity scale.

Table 2. *Questionnaire scales*

Scales	Total Item	Type of response to the item
Writing Essential (WE)	9	Strongly Disagree (1) to Strongly Agree (5)
Rational Reflective Writing (RRW)	8	Strongly Disagree (1) to Strongly Agree (5)
Creative Identity (CI)	8	Strongly Disagree (1) to Strongly Agree (5)
Technology Use to Support Writing & Writing Creativity (TUSW)	9	Strongly Disagree (1) to Strongly Agree (5)

Translation into Bahasa

The original questionnaires were written in English and translated into the local language (i.e., Bahasa Indonesia) in the current research in purpose to let the participants easily understand and relate to the questionnaires based on their experience. The translated questionnaires were proofread and validated by other researchers who were fluent both in English and Bahasa Indonesia (Mulyono et al., 2020; Ningsih et al., 2021).

Data analysis

The instruments of the research are gathered using Google Forms. The data recorded in the form is then saved in xlsx format (Microsoft Excel form). The time records are then omitted to retrieve only the data that is required. The participants' responses were then sorted, categorized, marked, and converted into the predetermined code on the following sheets. Before testing the psychometric properties of the scales, all the xlsx data was converted to formatted text (Mulyono et al., 2021) and underwent a two-phase evaluation process. The first phase was conducted to eliminate null responses and identify outliers, which would then be removed from the data (Goh et al., 2010). Data was identified as an outlier if they did not meet the criteria of an MNSQ >2. Of the 655 samples, 73 were found to be outliers, leaving 582 samples to be analyzed in the second phase. This phase was used to re-examine the respondents' data that had previously been reduced by the number of outliers (Linacre, 2010). Several analyses were done in this phase, including: (a) dimensionality; (b) reliability for the item and the person separation; (c) rating scale; (d) item bias; and (e) item-differential function (Lee et al., 2020; Mulyono et al., 2020, 2021; Ningsih et al., 2021; Zulaiha & Mulyono, 2020).

Findings and Discussion

The analysis of the unidimensionality of the items

Rasch unidimensionality analysis is known as the statistical analysis to determine whether an item is exclusively associated with one construct, and not compatible with any other construct (Yu, 2020). To examine the instrument unidimensionality, the Rasch Principal Component Analysis (PCA) was done to the global scale and its subscales (Mulyono et al., 2020). Table 4 presents the raw variance data which satisfies the criterion of

unidimensionality with figures of 20% or more. The global scale was measured at 58.5%, and subscales at Writing Essential (61.2%), Rational Reflective Writing (60.7%), Creative Identity (65.1%), and Technology Use to Support Writing & Writing Creativity (64.0%). The PCA eigenvalue for the first contrast was shown to be 3.2 logits for the global scale, 1.7 logits for the WE scale, 2.1 logits for the RRW scale, 1.6 logits for the CI scale, and 2.2 logits for the TUSW scale. These findings suggest that the global, RRW, and TUSW scales reflect multidimensionality or could incorporate other potential constructs; conversely, the other two scales (i.e., WE and CI) appear to be unidimensional.

Table 3. *Summary of Rasch measurement model on global scale and constructs*

Parameter (with quality criteria)	**Global scale (34 items)**	**WE (9 items)**	**RRW (8 items)**	**CI (8 items)**	**TUSW (9 items)**
Model fit: Summary of items					
Item mean in logit (criteria 0.0 logits)	.00, SD = 0.50	.00, SD = 0.50	.00, SD = 0.60	.00, SD = 0.23	.00, SD = 0.31
Item reliability	0.97	0.97	0.98	0.84	0.91
Item separation reliability (criteria: good, 0.81-0.90; very good, 0.91-0.94; excellent, >0.94)	0.97	0.97	0.98	0.84	0.91
Item model fit MNSQ range extremes (criteria: good. 0.5-1-5; very good, 0.71-1.4; excellent 0.77-1,3)	Infit: 0.72-1.50 Outfit: 0.68-1.47	Infit: 0.82-1.25 Outfit: 0.75-1.19	Infit: 0.83-1.30 Outfit: 0.78-1.35	Infit: 0.87-1.17 Outfit: 0.78-1.14	Infit: 0.76-1.15 Outfit: 0.70-1.09
Item separation index (criteria >3)	5.67 (6)	5.26 (3)	6.65 (5)	2.30 (2)	3.25 (3)
Separate item strata == [(4 x separation index) + 1]/3 (criteria: fair, 2-3; good, 3-4; very good, 4-5; excellent, >5)	7.89 = 8 level	7.35 = 7 level	9.20 = 9 level	3.40 = 3 level	4.66 = 5 level
Model fit: Summary of persons					
Person means in logits (criteria: 0.0 logits)	2.65 SD = 2.22	2.78 SD = 2.40	3.23 SD = 2.19	3.20 SD = 2.82	2.83 SD = 2.56
Person reliability	0.97	0.89	0.87	0.89	0.90
Person separation reliability (criteria: good, 0.81-0.90; very good, 0.91-0.94; excellent, >0.94)	0.97	0.89	0.87	0.89	0.90
Person separation index (criteria >2)	5.55 (6)	2.84 (3)	2.59 (3)	2.85 (3)	2.97 (3)
Separate Person strata = [(4 x separation index) + 1]/3 (criteria: fair, 2-3; good, 3-4; very good, 4-5; excellent, >5)	7.73 = 8 level	4.12 = 4 level	3.79 = 4 level	4.13 = 4 level	4.29 = 4 level

Table 4. *Rating scale and dimensionality of rasch measurement model on global scale and constructs*

Rating Scale Analysis					
Responses per category (criteria: >10)	**YES**	**NA**	**NA**	**NA**	**NA**
Adjacent threshold distance (criteria 1.4-5 logits)	(i) scale 0-1 was (0) - (-6.07) = 6.07 logits; (ii) scale 1-2 was (-6.07) – (-1.41) = 4.66 logits; (iii) scale 2-3 was (-1.41) – (1.60) = 3.01 logits; (iv) scale 3-4 was (1.60) – (5.88) = 4.28 logits				
Outfit MNSQ (Criteria: <2 logits)	YES	NA	NA	NA	NA
Probability curve graph (criteria: decent curve on each response category and each item is higher than 0.5 logits)	YES	NA	NA	NA	NA
Average measure (criteria: increase significantly overrating scale)	YES	NA	NA	NA	NA
Dimensionality					
Raw variance in data explained by measure (criteria: 20% fair, 40% very good, >60% excellent)	58.5%	61.26%	60.7%	65.1%	64.0.%
PCA eigenvalue for first contrast (criteria: >2.0 indicates the presence of another dimension, < 2 supports unidimensional scale)	3.2	1.7	2.1	1.6	2.2
Unexplained variance in 1st – 5th contrast of PCA of residuals (criteria: good, 5-10%; very good, 3-5%; excellent, <3%)	3.9% - 1.9%	7.4% - 4.5%	7.1% - 4.5%	7.0% - 4.6%	8.9% - 3.7%

SD: standard deviation; PCA: principal components analysis; DIF: differential item functioning; MNSQ: mean square; NA: not applicable. The table was adopted from the study by Sumintono et al., (Lee et al., 2020).

Item and person separation reliability

The criteria for item separation reliability range from good to excellent. Specifically, 0.81-0.90 is defined as good; 0.91-0.94 represents very good, and logits above 0.94 indicate excellent quality (Cohen et al., 2018). As seen in Table 3, the data shows that the global score reaches 0.97 logits and surpasses the criteria that reflect excellent quality. Furthermore, the constructs varied in quality from one another. In line with the global score, the WE and RRW scales both reflected excellent quality with logits of 0.97 and 0.98, respectively. However, the CI only met the criteria for good quality with 0.84, and the TUSW reached 0.91 logits and reflected very good quality. Furthermore, the criteria for person separation reliability also demonstrated a different situation, where the global score was the only one that met the excellent quality with 0.97 logits, whereas the WE, RRW, CI, and TUSW all represented good quality, with logits of 0.89, 0.87, 0.89, and 0.90, respectively. As a result, the consistency of respondents' responses to the data needed in this research was possible to be generalized from good to excellent. This clarifies that it is more practical to be able to define their self-efficacy in academic writing.

In addition, with the exception of the CI scale, the quality of the item separation reliability was excellent, as they were all higher than the expected criteria of >3. They were 5.67 logits on the global scale, 5.26 logits on the WE, 6.65 logits on the RRW, 2.30 logits on the CI, and 3.25 logits on the TUSW, which indicates the classification of the item difficulty. The person separation index on the global scale and each construct also presented excellent quality, meeting the criteria of >2, with 5.55 logits on the global scale, 2.84 logits on the WE, 2.59 logits on the RRW, 2.85 logits on the CI, and 2.97 logits on the TUSW. It can thus be concluded that the respondents showed excellent quality in answering the instrument.

Effectiveness of the rating scale

The questionnaire was developed using a 5-point Likert scale, with values ranging from 1 (strongly disagree) to 5 (strongly agree), for each item. The criteria for the response category were set at higher than 10. Data from the "observed persons" column showed that scale number one met the criteria of 10 or more persons with 655 persons voting on that scale. Additionally, the criteria for adjacent threshold distance had to be between 1.4 to 5 logits. It is shown in the data in Table 4, rating scale section, that the first scale received 6.07 logits, the second scale was 4.66 logits, the third scale was 3.01 logits, and the fourth scale was 4.28 logits. All of them have met the criteria.

Table 5. *Summary of category structure of rating scales*

Rating scale step number	Observed person	Average calibration	Outfit MNSQ	Threshold	Threshold distance
1	106 (1)	-5.58	1.46	NONE	(-7.18)
2	1190 (6)	-1.56	1.23	-6.07	-3.74
3	5017 (25)	1.25	.89	-1.41	.10
4	10525 (53)	3.32	.93	1.60	3.75
5	2950 (15)	5.41	1.03	5.88	(6.99)

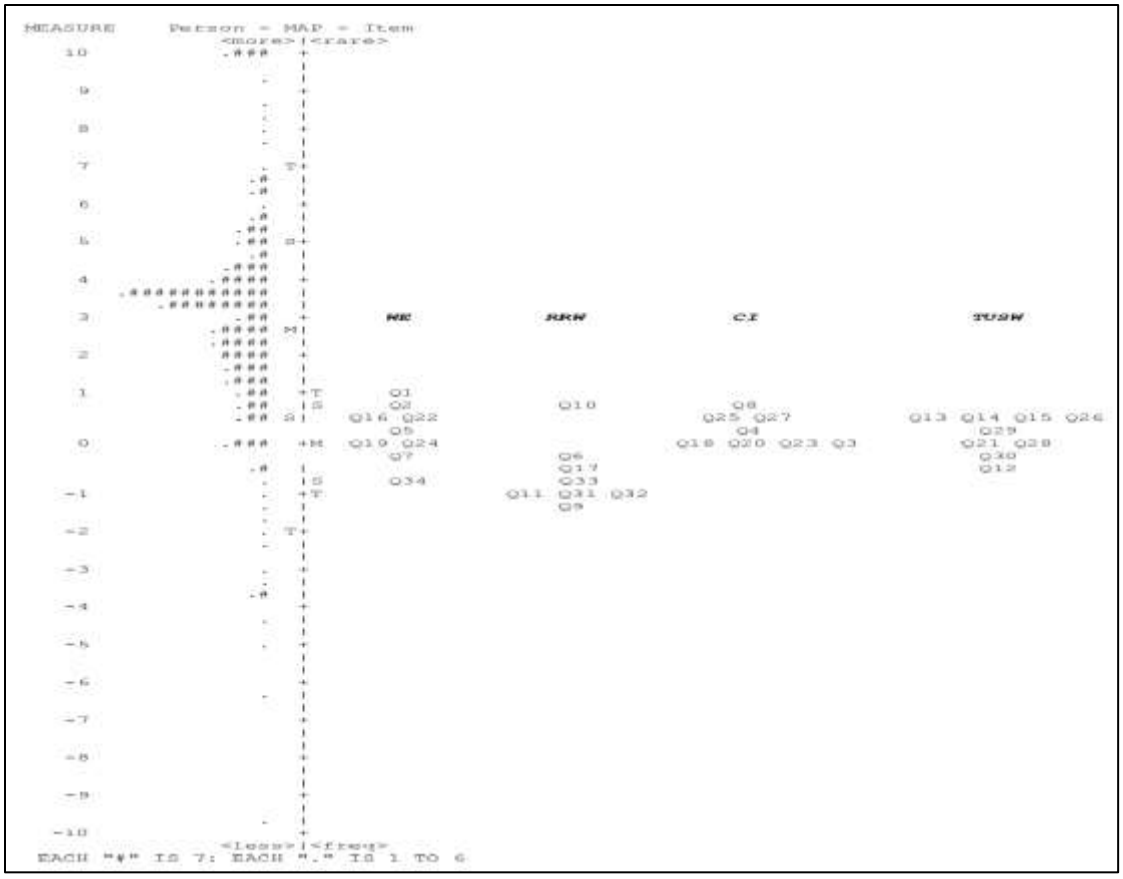
In addition, as shown in Table 5, the threshold distance was shown to increase for the rating scale step number one to five, beginning with (-7.18) to (6.99). The outfit MNSQ for each scale was shown lower than two logits and the curve measure of the response probability also met the requirement of > 0.5 logits. When the item is higher than 0.5 logits, it will cross the scale one another and shape a decent curve of the response category. Finally, the average measure denotes a significant and gradual rise of the average calibration from the negative pole to the positive pole: (-5.58) to (-1.56); (-1.56) to (1.25); (1.25) to (3.31); and (3.31) to (5.41).

Item and person mapping

Figure 1 shows the Wright Map, which represents the item- and person-mapping of the current study. The first line of the table provides information regarding the section distribution. The table is divided by a dotted line; the left side refers to the person items, while the right side describes the information on the item difficulty per construct. The least approved option is given the highest level while the most approved option is labeled the lowest level. From the figure, it is evident that Q1, Q2, Q10, and Q8 are reflected as the most difficult or rarely selected items by respondents for each scale are These items inquire about the ability of the respondents to generate ideas for academic writing, write texts that are clearly comprehensible, determine the expectations of readers, and develop impressive and professional writing. This suggests that most participants experienced difficulty in coming up with an idea in academic writing as they are unaware of the preferences of their readers and whether they can meet these expectations through their writing. In addition, as reflected in Q2 and Q8, most of the participants were unsure if they could write clear and impressive academic writing professionally which could be easily understood by the readers.

In addition, several items (i.e. Q16, Q22, Q25, Q27, Q13, Q14, Q15, and Q26) posit at the moderate level as in the map. The item Q16, Q22, Q25, Q27, Q13, Q14, Q15, and Q26 address the following questions: "Can I synthesize information from various sources of literature to build ideas and arguments in academic texts?"; "Can I paraphrase the original text accurately and write it in an academic format?"; "Can I vary the grammatical arrangement to maintain originality of meaning?"; "Am I able to identify incorrect grammar and rearrange it with more correct grammar?"; "Am I familiar with various websites or journal search engines and databases for references to support academic writing?"; "Can I use a digital reference management application to cite references in an academic text?"; "Can I save a list of articles that I cite into a database, and manage them using a digital reference management application?"; and "Can I use software or other digital applications to identify variations in grammar while still maintaining the originality of meaning?" respectively.

Figure 1. *Item and person mapping*



"#" represents seven persons while "." represents one to six persons. M_p: person means; S_p: one standard deviation of a person means; T_p: two standard deviations of a person mean; M_i: item mean; S_i: one standard deviation of item means; T_i: two standard deviations of item mean. Furthermore, as shown in Figure 1 above, the easiest level of items answered by the participants were found to be Q11, Q31, Q32, and Q9. These items were found at or below the line of "t" which represents the lower level of the item mapping. The items address the questions "Do I always re-read academic writing and reflect on its content in order to write better texts?", "Do I have an open mind to receive feedback from colleagues regarding the content of the academic texts I write?", "Do I constantly seek feedback from my colleagues by asking them to read and comment on my work?", and "When writing, do I always consider whether the information I have written is clear and understandable for the reader?". In summary, it can be concluded that most of the participants demonstrate consideration when writing academically, and have the willingness to improve by anticipating feedback and advice from their colleagues in order to write better and more legible academic texts. They also have consideration for the readers of their writing, double checking their work to ensure that the information they have written is comprehensible for readers.

Item bias distributions

Person - Differential Item Functioning (DIF) analysis was done to determine which items on which instrument are biased or benefited one particular party in demographic data. The analysis threshold for DIF analysis includes the DIF contrast value should be higher than 0.5 and the probability value is lower than 0.05 (Chan & Subramaniam, 2020; J M Linacre, 2011; Mulyono et al., 2020). The results, shown in Table 6 and Table 7, indicated biased items for two out of the three demographics (i.e., status and background of study). Table 6 displays the biased items for the participant status demographic, while Table 7 displays the biased items for the background of education demographic. Fortunately, no biased items were detected in the gender demographic. Further analysis of demographic and construct items which did not meet the criteria was conducted by taking into consideration the construct item, the type of demographic, and the DIF measure.

Table 6. *Item bias demography status*

No	Item	Status	DIF Measure	DIF Contrast	t	Probability
1	Q4	H	.57	.64	2.54	.0122
		I	-.07			
2	Q8	J	.93	.55	3.05	.0024
		H	.38			
3	Q10	J	1.00	.76	4.20	.0000
		H	.24			
4	Q14	I	1.13	1.18	4.80	.0000
		H	-.06			
		J	.58	.64	3.50	.0005
		H	-.06			
5	Q15	I	.76	.63	2.56	.0116
		H	.13			
		J	.73	.59	3.25	.0012
		H	.13			
6	Q31	H	-.48	1.41	5.25	.0000
		I	-1.88			
		H	-.48	.72	3.80	.0002
		J	-1.20			
		J	-1.20	.68	2.49	.0139
		I	-1.88			
7	Q32	H	-.46	1.04	3.92	.0001
		I	-1.50			
		H	-.46	.60	3.15	.0017
		J	-1.06			
8	Q33	H	-.48	.60	2.28	.0242
		I	-1.07			

Note. H = pre-service teacher; I = TVET instructor; J = school teacher

As seen in Table 6, several bias items were identified in participants' status demography. Firstly, in Q4, it was found that pre-service teacher respondents were more capable of expressing their individual ideas and opinions when writing an academic text than

TVET instructors. Secondly, Q8 revealed that school teacher respondents were more adept at writing professional and impressive academic texts than pre-service respondents. Additionally, Q10 suggested that school teacher respondents were more capable of understanding what the reader wanted and meeting those expectations in their writing than pre-service teachers. Furthermore, Q14 found that pre-service teachers had a lesser understanding of utilizing digital reference management applications to cite references in academic texts than the other two groups. Lastly, Q15 also demonstrated that pre-service teachers had a weaker understanding of developing storage for a database of cited articles by managing them with digital reference management applications than the other two groups.

However, it was demonstrated in Q31 that pre-service teacher respondents were more open-minded to receiving feedback from colleagues about the content of academic texts they had written than the other two groups. Furthermore, it was revealed in Q32 that the pre-service teachers were more likely to ask colleagues for feedback after reading their academic text, compared to the other two groups. Finally, Q33 indicated that pre-service teacher respondents were more likely to accept or reject feedback provided by their colleagues on written revisions, compared to the TVET instructor respondents.

Table 7. *Item bias demography background of study*

No	Item	Status	DIF Measure	DIF Contrast	t	Probability
1	Q10	Y	1.04	.80	2.34	.0236
		V	.24			
		X	.89	.65	3.71	.0002
		V	.24			
2	Q12	V	-.30	.80	2.17	.0350
		Y	-1.10			
		W	.64	1.74	2.29	.0426
		Y	-1.10			
3	Q13	V	.10	.84	2.28	.0270
		Y	-.74			
		W	1.50	2.24	3.05	.0110
		Y	-.74			
		X	.38	1.12	3.04	.0039
		Y	-.74			
4	Q14	X	.71	.77	4.37	.0000
		V	-.06			
		Y	.73	.79	2.27	.0274
		V	-.06			
5	Q15	X	.76	.62	3.56	.0004
		V	.13			
6	Q31	V	-.48	1.99	2.60	.0356
		W	-2.47			
		V	-.48	.86	4.70	.0000
		X	-1.33			
		V	-.48	.99	2.67	.0105
		Y	-1.46			
7	Q32	V	-.46	.69	3.77	.0002
		X	-1.15			

Note. V = university students; W = diploma degree; X = bachelor degree; Y = master degree; Z = doctoral degree

As shown in Table 7, other bias items were found in participants' education backgrounds. In Q10, it was demonstrated that participants who are still university students had less capability in determining what readers want to know from their writing and how to meet those expectations than the respondents with Bachelor's and Master's degrees. Additionally, Q12 revealed that those holding a Master's degree had a reduced understanding of how to find literature sources from various journals, repositories, or other digital databases, in comparison to those who were still university students or had attained a Diploma degree. Similarly, in Q13, the respondents with a Master's degree displayed less awareness than the other three groups (university students, Diploma holders, and Bachelor's degree holders) for various websites, journal search engines, or databases for referencing to support their academic writing.

Furthermore, in Q14, it was found that university student respondents had less understanding of how to use digital reference management applications to cite references in academic texts than the respondents with bachelor's or master's degrees. Similarly, in Q15, it was found that respondents who had graduated from graduate school had a greater understanding of how to use digital references to develop a database of articles they had cited than those who were still university students. Conversely, in Q31, it was found that university student respondents were more open to receiving feedback from colleagues on the content of academic texts they wrote than the other three groups (i.e., respondents with diplomas, bachelors, and master's degrees). Finally, according to Q32, university student respondents were more willing to seek out feedback by asking for their colleagues' suggestions after reading their academic texts than respondents with bachelor's degrees.

Conclusion

Despite the number of current studies on academic writing self-efficacy offering results of investigation of the variable, there is rarely an examination of the instrument measuring academic writing creativity and self-efficacy, specifically in Indonesian writing teachers contexts. The current study examined the psychometric characteristic of the academic writing creativity and self-efficacy instruments adapted and expanded from the earlier study by Mitchell (2021) measuring the level of academic writing creativity and self-efficacy, and the effect of five affective variables (i.e., writing essential scale, rational reflective writing scale, creativity identity scale, and technology used to support writing and writing creativity scale). Despite the high level of reliability of the academic writing creativity and self-efficacy scale (Cronbach's $\alpha > 0.95$ for the global scale, and Cronbach's $\alpha > 0.85$ for each constructs) and sufficient separation index for the person and item aspects, the instruments present an issue concerning the multidimensionality i.e. on the global scale, Rational Reflective Writing (RRW), and Technology Use to Support Writing & Writing Creativity (TUSW); indicating that the questionnaire could incorporate other potential constructs on the related aspects. The analysis of DIF resulted that the academic writing creativity and self-efficacy scale could potentially be biased when employed to collect the data from participants with differing teaching statuses and educational backgrounds. These findings thus suggest that caution should be taken when using the Academic Writing Creativity and Self-Efficacy Scale, as highlighted by the results of the current study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest.

Funding

The current study was funded by the directorate general of higher education, research and technology, Ministry of education, culture, research and technology, Republic of Indonesia, through a research grant PDUPT year 2022.

References

Alberth. (2019). Use of Facebook, students' intrinsic motivation to study writing, writing self-efficacy and writing performance. *Technology, Pedagogy and Education*, *28*(1), 21–36. https://doi.org/10.1080/1475939X.2018.1552892

Allison, D. (2004). Creativity, students' academic writing, and EAP: exploring comments on writing in an English language degree programme. *Journal of English for Academic Purposes*, *3*(3), 191–209. https://doi.org/https://doi.org/10.1016/j.jeap.2003.11.005

Badley, G. F. (2019). Post-Academic Writing: Human Writing for Human Readers. *Qualitative Inquiry*, *25*(2), 180–191. https://doi.org/10.1177/1077800417736334

Bandura, A. (1978). Self-efficacy: Toward a unifying theory of behavioral change. *Advances in Behaviour Research and Therapy*, *1*(4). https://doi.org/10.1016/0146-6402(78)90002-4

Bandura, A. (1994). *Self-Efficacy. In V. S. Ramachaudran (Ed.), Encyclopedia of Human Behavior (Vol. 4, pp. 71-81).*

Bandura, A., & Adams, N. E. (1977). Analysis of self-efficacy theory of behavioral change. *Cognitive Therapy and Research*, *1*(4). https://doi.org/10.1007/BF01663995

Bandura, A., Freeman, W. H., & Lightsey, R. (1997). Self-efficacy: The exercise of control. In *Journal of Cognitive Psychotherapy* (Vol. 13, Issue 2, pp. 158–166). https://doi.org/10.1891/0889-8391.13.2.158

Callinan, C. J., van der Zee, E., & Wilson, G. (2018). Developing essay writing skills: an evaluation of the modelling behaviour method and the influence of student self-efficacy. *Journal of Further and Higher Education*, *42*(5), 608–622. https://doi.org/10.1080/0309877X.2017.1302564

Chan, M., & Subramaniam, R. (2020). Validation of a science concept inventory by Rasch analysis. In M. S. Khine (Ed.), *Rasch Measurement: Applications in Quantitative Educational Research* (Issue July, pp. 159–178). Springer Singapore. https://doi.org/10.1007/978-981-15-1800-3_9

Chen, J., & Zhang, L. J. (2019). Assessing student-writers' self-efficacy beliefs about text revision in EFL writing. *Assessing Writing*, *40*, 27–41. https://doi.org/10.1016/j.asw.2019.03.002

Cohen, L., Manion, L., Morrison, K., & Bell, R. (2018). *Research methods in education* (8th ed.). Routledge.

Collado, S., Fidalgo, C., Rodríguez-Rey, R., & Sorrel, M. A. (2023). Development and Validation of the Self-efficacy for Writing and Defending Academic Texts Scale1. *Psicología Educativa*, *29*(1), 91–99. https://doi.org/10.5093/psed2022a15

Cremin, T., & Oliver, L. (2017). Teachers as writers: a systematic review. *Research Papers in Education*, *32*(3), 269–295. https://doi.org/10.1080/02671522.2016.1187664

Dobele, A. R., & Veer, E. (2019). My best writing space: understanding academics self-professed writing spaces. *Higher Education*, *78*(2), 345–364. https://doi.org/10.1007/s10734-018-0346-y

Goh, J. W. P., Quek, C. J., & Lee, O. K. (2010). An Investigation of students' perceptions of learning benefits of weblogs in an east Asian context: A Rasch analysis. *Journal of Educational Technology & Society*, *13*(2), 90–101. http://www.jstor.org/stable/jeductechsoci.13.2.90

Groot, B. C., Vink, M., Haveman, A., Huberts, M., Schout, G., & Abma, T. A. (2019). Ethics of care in participatory health research: mutual responsibility in collaboration with co-researchers. *Educational Action Research*, *27*(2), 286–302. https://doi.org/10.1080/09650792.2018.1450771

Huerta, M., Goodson, P., Beigi, M., & Chlup, D. (2017). Graduate students as academic writers: writing anxiety, self-efficacy and emotional intelligence. *Higher Education Research & Development*, *36*(4), 716–729.

Hyland, K. (2013). Writing in the university: Education, knowledge and reputation. *Language Teaching*, *46*(1), 53–70. https://doi.org/10.1017/S0261444811000036

Joshi, S. C., Naik, J., & Aggarwal, D. (2022). Formalizing Creativity: Academic Writing in Early College Years. *Indian Journal of Mass Communication and Journalism*, *1*(3), 25–28. https://doi.org/10.54105/ijmcj.C1011.031322

Kiriakos, C. M., & Tienari, J. (2018). Academic writing as love. *Management Learning*, *49*(3), 263–277. https://doi.org/10.1177/1350507617753560

Lee, W. L., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A Rasch analysis. *European Journal of Preventive Cardiology*, *28*(12), e1-e5. https://doi.org/10.1177/2047487320902322

Linacre, J M. (2011). A user's guide to WINSTEPS [Computer Manual]. *Chicago: Winsteps*.

Linacre, John M. (2010). When to stop removing items and persons in Rasch misfit analysis? *Rasch Measurement Transactions*, *23*(4), 1241. https://www.rasch.org/rmt/%0Armt234g.htm

Mccarthy, L. P. (1987). A Stranger in Strange Lands: A College Student Writing across the Curriculum. In *Source: Research in the Teaching of English* (Vol. 21, Issue 3).

Meza, P., & González, M. (2020). Construction and validation of the self-efficacy scale for disciplinary academic writing. *Cogent Education*, *7*(1). https://doi.org/10.1080/2331186X.2020.1830464

Mickwitz, Å., & Suojala, M. (2020). Learner autonomy, self-regulation skills and self-efficacy beliefs-How can students' academic writing skills be supported? *Language Learning in Higher Education*, *10*(2), 381–402. https://doi.org/10.1515/cercles-2020-2026

Mitchell, K. M., & McMillan, D. E. (2018). A curriculum-wide assessment of writing self-efficacy in a baccalaureate nursing program. *Nurse Education Today*, *70*, 20–27. https://doi.org/10.1016/j.nedt.2018.08.003

Mitchell, K. M., McMillan, D. E., Lobchuk, M. M., Nickel, N. C., Rabbani, R., & Li, J. (2021). Development and validation of the Situated Academic Writing Self-Efficacy Scale (SAWSES). *Assessing Writing*, *48*, 100524. https://doi.org/https://doi.org/10.1016/j.asw.2021.100524

Mulyono, H., Saskia, R., Arrummaiza, V. S., & Suryoputro, G. (2020). Psychometric assessment of an instrument evaluating the effects of affective variables on students' WTC in face-to-face and digital environment. *Cogent Psychology*, *7*(1), 1823617. https://doi.org/https://doi.org/10.1080/23311908.2020.1823617

Mulyono, H., Suryoputro, G., & Jamil, S. R. (2021). The application of WhatsApp to support online learning during the COVID-19 pandemic in Indonesia. *Heliyon*, *7*(8), e07853. https://doi.org/https://doi.org/10.1016/j.heliyon.2021.e07853

Murray, R. (2013). *Writing for Academic Journals* (3rd ed.). Open University Press.

Ningsih, S. K., Mulyono, H., Ar Rahmah, R., & Fitriani, N. A. (2021). A Rasch-based validation of EFL teachers' received online social support scale. *Cogent Education*, *8*(1), 1957529. https://doi.org/https://doi.org/10.1080/2331186X.2021.1957529

Odena, O., & Burgess, H. (2015). How doctoral students and graduates describe facilitating experiences and strategies for their thesis writing learning process: a qualitative approach. *Studies in Higher Education*, *42*(3), 572–590. https://doi.org/10.1080/03075079.2015.1063598

Pajares, F. (2003). Self-efficay beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, *19*(2), 139–158. https://doi.org/10.1080/10573560308222

Plakhotnik, M. S., & Rocco, T. S. (2016). Increasing Writing Self-Efficacy of Adult Learners: Different Approaches, Different Results. *Adult Learning*, *27*(4), 160–167. https://doi.org/10.1177/1045159515611879

Ruegg, R. (2018). The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy. *Language Learning Journal*, *46*(2), 87–102. https://doi.org/10.1080/09571736.2014.958190

Teng, L. S., Sun, P. P., & Xu, L. (2018). Conceptualizing Writing Self-Efficacy in English as a Foreign Language Contexts: Scale Validation Through Structural Equation Modeling. *TESOL Quarterly*, *52*(4), 911–942. https://doi.org/10.1002/tesq.432

Van de Poel, K., & Gasiorek, J. (2012). Effects of an efficacy-focused approach to academic writing on students' perceptions of themselves as writers. *Journal of English for Academic Purposes*, *11*(4), 294–303. https://doi.org/10.1016/j.jeap.2012.07.003

Wei, W., Cheong, C. M., Zhu, X., & Lu, Q. (2022). Comparing self-reflection and peer feedback practices in an academic writing task: a student self-efficacy perspective. *Teaching in Higher Education*. https://doi.org/10.1080/13562517.2022.2042242

Wingate, U., & Tribble, C. (2012). The best of both worlds? Towards an English for Academic Purposes/Academic Literacies writing pedagogy. *Studies in Higher Education*, *37*(4), 481–495. https://doi.org/10.1080/03075079.2010.525630

Yu, C. H. (2020). Objective measurement: How Rasch modeling can simplify and enhance your assessment. In M. Khine (Ed.), *Rasch Measurement: Applications in Quantitative Educational Research* (pp. 47–73). Springer. https://doi.org/10.1007/978-981-15-1800-3_4

Zulaiha, S., & Mulyono, H. (2020). Exploring junior high school EFL teachers' training needs of assessment literacy. *Cogent Education*, *7*(1), 1772943. https://doi.org/10.1080/2331186X.2020.1772943

Zumbrunn, S., Broda, M., Varier, D., & Conklin, S. (2020). Examining the multidimensional role of self-efficacy for writing on student writing self-regulation and grades in elementary and high school. *British Journal of Educational Psychology*, *90*(3), 580–603. https://doi.org/10.1111/bjep.12315

Biographical Notes

HERRI MULYONO is an associate professor at University of Muhammadiyah Prof. DR. HAMKA, Indonesia

SRI KUSUMA NINGSIH is a lecturer at University of Muhammadiyah Prof. DR. HAMKA, Indonesia

BEAUTY SHOLEHA RAUFI was a student at University of Muhammadiyah Prof. DR. HAMKA, Indonesia