Person Fit Analysis for Assessing Academic Writing Performance Using Rasch Model

Yenni Arif Rahman[1], Leny Hikmah Rentiana [2], Ulfa Rahma Dhini [3]
[1,2,3] Universitas Bina Sarana Informatika
Email: yeni.yar@bsi.ac.id.

Journal info
Glasser Education Journal
p-ISSN: 2579-5082
e-ISSN: 2598-2818
DOI: 10.32529/glasser.v7i2.2571
Volumes: 7
Number: 2
Month: 2023

Keywords: *Person Fit, Writing Assessment, Rasch Model*

Abstract.
Assessing academic writing performance is a critical aspect of educational practices. The thesis is derived from writing is an essential skill that students need to develop to succeed academically and professionally. To fulfill this, the Rasch model offers important test statistics that play a crucial role in constructing tests related to evaluation and item selection problems, as well as in decision-making regarding the generated test scores. Building upon this narrative, this research is conducted with the aim of identifying person fit in assessing academic writing performance using the Rasch model. The sample in the form of essays was obtained from a group of 40 students who had previously undergone a six-month academic writing program. The data was then analyzed using Ministep 4.8.2.0 built for Rasch Model analysis. The analysis results indicated that there were 31 students in the person misfit category, suggesting a need to reconsider the appropriateness of the conducted treatment and whether there were other factors contributing to it. Additionally, a discrepancy was found in 5 students with persons misfit or overfit, and their better performance was observed on more difficult items, contrary to the Rasch model's profile.

A. INTRODUCTION

Assessing academic writing performance is a critical aspect of educational practices. The thesis is derived from writing is an essential skill that students need to develop to succeed academically and professionally. To do that academic writing should demonstrate knowledge, present arguments, and communicate ideas effectively. Therefore, it is important to evaluate students' writing skills accurately and reliably to assess whether those qualities have been accommodated properly and adequately in students' essays.

Traditional assessment methods, such as rubrics and checklists, may not provide the rigor required for reliable assessments (Fisher, 2007; Wibisono, 2018). These methods rely on subjective judgments by evaluators, which can lead to inconsistent and unreliable results. As a result, many researchers have explored statistical methods to assess writing performance objectively and reliably (Bond & Fox, 2007; Boone & Yale, 2014). It is a statistical method that has been widely used in educational research to evaluate various constructs, including academic writing performance (Engelhard, 1992; Fisher, 2007;

Razak & Thien, 2012). It also provides an objective and reliable method for evaluating writing performance. It allows for the analysis of item and participant responses to evaluate writing skills accurately.

In recent years, there has been a growing interest in the application of the Rasch Model to assess person fit in academic writing performance. Several studies have investigated the use of this model to evaluate writing skills and identify students who may require additional support or challenge to improve their performance. A recent study by Osman et al (2012) examined the use of the Rasch Model to assess person fit into a group of Chinese graduate students' academic writing performance. The study found that the Rasch Model provided a reliable and valid means of assessing writing skills and identifying students who required additional support. Another study by Li et al (2021) investigated the use of the Rasch Model to assess the writing skills of Chinese undergraduate students. The study found that person fit analysis using the Rasch Model could accurately identify students who were performing below their expected ability level and who required additional support to improve their writing skills. A study by Rahman (2023) examined the application of the Rasch Model to assess person fit and item fit in a group of Indonesian students' academic writing performance in blended learning program. The study found that person fit analysis using the Rasch Model could accurately identify students who were performing above their expected ability level and who required more challenging writing tasks to further develop their writing skills.

Despite the Rasch Model's usefulness in evaluating writing performance, limited research has explored its application in assessing person fit in academic writing assessment. Person fit refers to the degree to which an individual's responses fit with the Rasch Model. Identifying person fit issues is crucial as it can help identify students who have either cheated during the test or have lower or higher abilities than the Rasch Model estimated. Detecting person fit issues and taking appropriate actions can improve the accuracy and reliability of the assessment results.

In the context of academic writing performance, person fit analysis is essential as it can help educators and researchers identify students who require additional support to improve their writing skills. It can also help identify students who may be performing above their estimated ability, and thus, require more challenging writing tasks. Furthermore, person fit analysis can provide insights into the effectiveness of educational interventions aimed at improving students' writing skills. By identifying students with person fit issues, educators and researchers can monitor the effectiveness of educational interventions and make necessary adjustments to improve students' writing performance.

Therefore, this study aims to explore the application of the Rasch Model to analyze person fit in assessing academic writing

performance. The study's findings can contribute to the development of more objective and reliable methods for assessing academic writing performance. Moreover, the results can inform educational practices and the development of more effective interventions to support students' academic writing performance.

This study's significance lies in its potential to improve the accuracy and reliability of the assessment results. By analyzing person fit, educators and researchers can identify students who require additional support to improve their writing skills, which can enhance their academic and professional success. Additionally, the study's results can contribute to the development of more effective educational interventions to improve students' writing skills.

In conclusion, person fit analysis using the Rasch Model can provide insights into students' academic writing performance and inform educational practices to support their learning. The study's results can contribute to the development of more reliable and objective methods for assessing writing performance, leading to more effective interventions to improve students' writing skills.

B. RESEARCH METHODS

This study involved 40 EFL students from Eloquensi English Language Centre who had completed the TOEFL iBT essay writing course. They possessed an intermediate or advanced level of English proficiency. The main task assigned to them was to compose essays consisting of five paragraphs. The essays followed a standardized format with an introduction, content, and conclusion. The intention behind implementing this standardized structure was to guarantee impartial evaluations by the assessors, as the essay's paragraph count could potentially impact the rating, positively or negatively.

The Rasch model is employed in this study for analysis purposes due to its ability to capture the interaction between respondents and items simultaneously. In contrast to raw scores, the Rasch model employs logit values to express the probability of an item being chosen by a group of participants (Sumintono, B. & Widhiarso, 2013; Tan, 2013). The purpose of using the Rasch model is to estimate the expected raw score for Likert ratings, which are ordinal in nature and do not have equal intervals between scores. Andrich (1978) expanded the application of the Rasch model to polytomous data by incorporating two core principles: the measurement of individual ability or agreement level and the measurement of item difficulty in achieving agreement (Misbach & Sumintono, 2014). For data analysis purposes, the output utilized includes summary statistics (Figure 1) to gather reliability information. Additionally, the output includes unidimensionality items (Figure 2) and Fit Order items (Figure 3) for assessing validity.

This study employed a holistic rubric developed by Jacob et al (Jacobs et al, 1981) as the measurement tool. The rubric consists of six levels of measurement, namely proficient,

fluent, expanding, developing, beginning, and emerging. The criteria column reveals that the assessment focuses on four key elements of writing ability: content, structure, diction, and mechanics. Content encompasses the introduction, ideas or body paragraphs, and the logical organization of thoughts. Structure evaluates not only grammatical proficiency within sentences but also the composition of paragraphs using different sentence types (simple, compound, complex, and compound-complex). Diction assesses the respondent's vocabulary usage and word variations within a paragraph to avoid word repetition. Lastly, writing mechanics examines the correct implementation of punctuation, spelling, and capitalization.

Table 1. Jacob et al Holistic Rubric (1981)

Rating	Criteria
Proficient	1. Writes single or multiple paragraphs with clear introduction, fully develop idea, present idea logically 2. Uses appropriate verb tense and a variety of grammatical and syntactical structures; uses complex sentences effectively; uses smooth transitions 3. Uses varied, precise vocabulary 4. Has occasional errors in mechanics (spelling, punctuation, and capitalization) which do not detract from meaning
Fluent	1. Writes single or multiple paragraphs with main idea and supporting detail, present idea logically, though some parts may not fully develop 2. Uses appropriate verb tense and a variety of grammatical and syntactical structures; errors in sentence do not detract from meaning; uses transitions 3. Uses varied vocabulary appropriate for the purpose 4. Has few errors in mechanics which do not detract from meaning
Expanding	1. Organizes ideas in logical or sequential order with some supporting detail; begins to write a paragraph 2. Experiment with a variety of verb tenses, but does not use them consistently; subject/verb agreement errors; uses some compound and complex sentences; limited use of transitions 3. Vocabulary is appropriate to purpose but sometimes awkward 4. Use punctuation, capitalization, and mostly conventional spelling; errors sometimes interfere with meaning
Developing	1. Writes sentences around an idea; some sequencing present, but may lack of cohesion 2. Write in present tense and simple sentences; has difficulty with subject/verb agreement, run-on sentences are common; begin to use compound sentences 3. Uses high frequency words; may have difficulty with word order; omit endings or words 4. Uses some capitalization, punctuation and transitional spelling; errors often interfere with meaning
Beginning	1. Begin to convey meaning through writing 2. Write predominantly phrases and patterned or simple sentences 3. Uses limited or repetitious vocabulary 4. Uses temporary (phonetic) spelling
Emerging	1. No evidence of idea development or organization 2. Uses single word, pictures, and patterned phases 3. Copies from model 4. Little awareness of spelling, capitalization, or punctuation

The six measurement levels mentioned earlier are converted into five Likert ratings, as shown in table 2. It is necessary to interpret the values into a Likert scale in order to facilitate the further processing of the raw scores using ministep software.

Table 2. Rubric Rating Scale

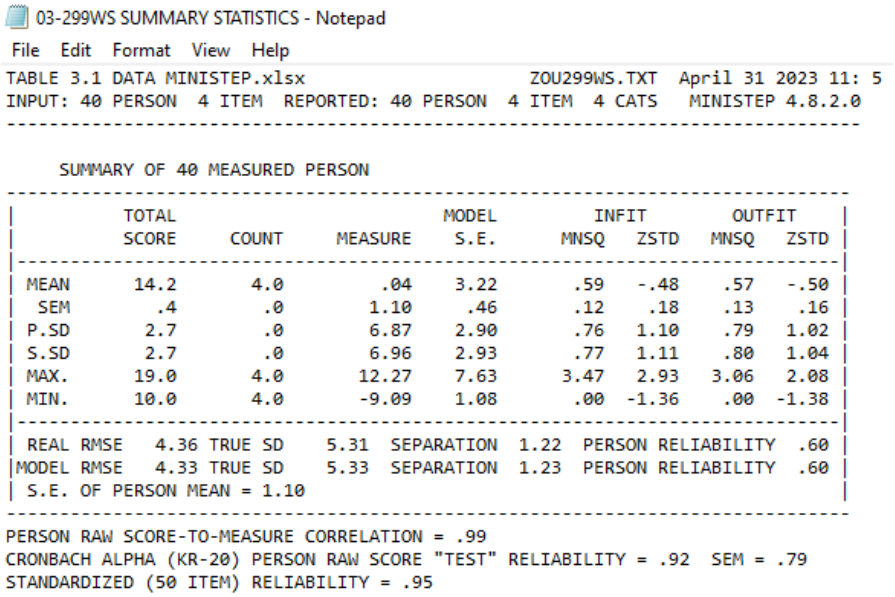
Scale	*Likert* Score
Proficient	5
Fluent	4
Expanding	3
Developing	2
Emerging & Beginning	1

Based on the necessary analysis, there are eight steps for analyzing person fit using the Rasch model.

1. Collect data on academic writing ability assessments. This data can be in the form of a scale or numerical values given by assessors for each item on the academic writing ability test.
2. Prepare the data in the appropriate format for Rasch analysis. The data must be in matrix form, with rows representing participants and columns representing test items.
3. Run the Rasch analysis on the data using statistical analysis software that allows for Rasch analysis. The results of the analysis will include calculations of participant ability values and item difficulty levels.
4. Conduct person fit analysis by examining participants' standard residuals. Standard residuals are the difference between the estimated ability of participants and their observed ability. High standard residuals indicate that participants have values that do not fit the Rasch model, indicating cheating in test answers.
5. Identify participants with high standard residuals and review their test answer results. It is possible that some items do not fit the participant's ability, or the participant may have cheated in answering the test.
6. Re-evaluate test items and participant values. Identify and eliminate test items that do not fit the Rasch model and/or participants who cheated in answering the test.
7. Perform a Rasch model analysis again on the updated data to check for improvements in person fit and suitability of data to the Rasch model.
8. Interpret the results of the Rasch and person fit analysis. These results can be used to inform practices in evaluating academic writing ability and improving the quality of academic writing ability tests and measurements.

C. RESULTS AND DISCUSSION

Three sets of output data are utilized to assess the validity and reliability of individuals and items in student essays. The initial set of output data comprises summary statistics. The second set involves item statistics, which aids in identifying items that do not fit well. Lastly, person statistics are examined to identify individuals who do not fit well. The data outputs for this study were obtained through the application of the Rasch model analysis using ministep software. This particular software is specifically designed for statistical analysis related to Rasch modeling



03-299WS SUMMARY STATISTICS - Notepad

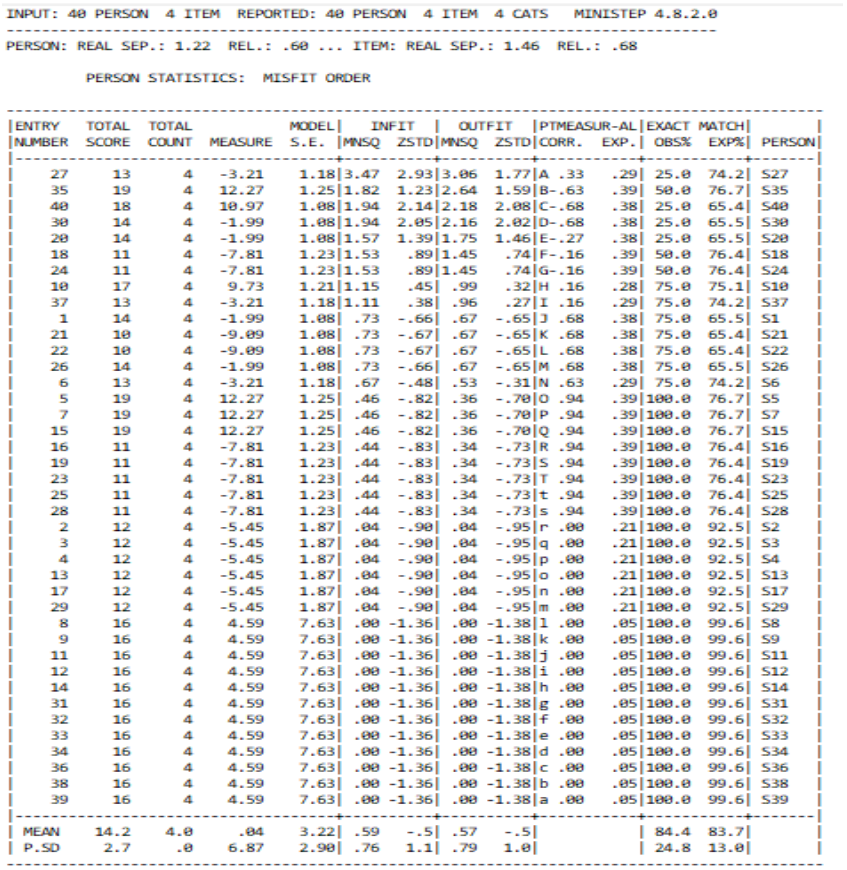
File Edit Format View Help

```
TABLE 3.1 DATA MINISTEP.xlsx                    ZOU299WS.TXT  April 31 2023 11: 5
INPUT: 40 PERSON  4 ITEM  REPORTED: 40 PERSON  4 ITEM  4 CATS   MINISTEP 4.8.2.0
-------------------------------------------------------------------------------

     SUMMARY OF 40 MEASURED PERSON
-------------------------------------------------------------------------------
|          TOTAL                        MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      14.2       4.0         .04    3.22       .59   -.48    .57   -.50 |
|  SEM        .4        .0        1.10     .46       .12    .18    .13    .16 |
| P.SD       2.7        .0        6.87    2.90       .76   1.10    .79   1.02 |
| S.SD       2.7        .0        6.96    2.93       .77   1.11    .80   1.04 |
| MAX.      19.0       4.0       12.27    7.63      3.47   2.93   3.06   2.08 |
| MIN.      10.0       4.0       -9.09    1.08       .00  -1.36    .00  -1.38 |
|-----------------------------------------------------------------------------|
| REAL RMSE   4.36 TRUE SD    5.31  SEPARATION  1.22  PERSON RELIABILITY  .60 |
|MODEL RMSE   4.33 TRUE SD    5.33  SEPARATION  1.23  PERSON RELIABILITY  .60 |
| S.E. OF PERSON MEAN = 1.10                                                  |
-------------------------------------------------------------------------------
PERSON RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .92  SEM = .79
STANDARDIZED (50 ITEM) RELIABILITY = .95
```

FIGURE 1. Summary Statistics

In this section, the summarized research findings are presented, as shown in Figure 1. Figure 1 displays the outfit MNSQ (outlier-sensitive or information-weighted fit Mean Square) statistics used to identify whether individuals fit or misfit the Rasch model, along with their accompanying characteristics. Figure 1 presents a summary of statistics. The person measure is 0.04, indicating the average value of respondents' performance in the essay writing assessment. An average value higher than 0.0 logit suggests that respondents tend to meet the standards outlined in the rubric. Cronbach's alpha value is utilized to assess the overall reliability of the interaction between individuals and items. The summary statistics reveal a Cronbach's alpha value of 0.92, indicating a high level of reliability. Person reliability demonstrates the consistency of measurements, indicating that repeated measurements will yield similar information. In other words, if another party were to conduct the same measurement, the values obtained would not differ significantly. Minor differences may still exist, which are considered acceptable. However, if significant differences arise in the results of the same sample analyzed by different researchers, several factors should be examined, including temporal similarity (stability), equivalence of assessment instruments, internal consistency of elements within the instrument, and agreement among raters. In Figure 1, the person reliability value is 0.60.



INPUT: 40 PERSON 4 ITEM REPORTED: 40 PERSON 4 ITEM 4 CATS MINISTEP 4.8.2.0

PERSON: REAL SEP.: 1.22 REL.: .60 ... ITEM: REAL SEP.: 1.46 REL.: .68

PERSON STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	PERSON
27	13	4	-3.21	1.18	3.47	2.93	3.06	1.77	A .33	.29	25.0	74.2	S27
35	19	4	12.27	1.25	1.82	1.23	2.64	1.59	B-.63	.39	50.0	76.7	S35
40	18	4	10.97	1.08	1.94	2.14	2.18	2.08	C-.68	.38	25.0	65.4	S40
30	14	4	-1.99	1.08	1.94	2.05	2.16	2.02	D-.68	.38	25.0	65.5	S30
20	14	4	-1.99	1.08	1.57	1.39	1.75	1.46	E-.27	.38	25.0	65.5	S20
18	11	4	-7.81	1.23	1.53	.89	1.45	.74	F-.16	.39	50.0	76.4	S18
24	11	4	-7.81	1.23	1.53	.89	1.45	.74	G-.16	.39	50.0	76.4	S24
10	17	4	9.73	1.21	1.15	.45	.99	.32	H .16	.28	75.0	75.1	S10
37	13	4	-3.21	1.18	1.11	.38	.96	.27	I .16	.29	75.0	74.2	S37
1	14	4	-1.99	1.08	.73	-.66	.67	-.65	J .68	.38	75.0	65.5	S1
21	10	4	-9.09	1.08	.73	-.67	.67	-.65	K .68	.38	75.0	65.4	S21
22	10	4	-9.09	1.08	.73	-.67	.67	-.65	L .68	.38	75.0	65.4	S22
26	14	4	-1.99	1.08	.73	-.66	.67	-.65	M .68	.38	75.0	65.5	S26
6	13	4	-3.21	1.18	.67	-.48	.53	-.31	N .63	.29	75.0	74.2	S6
5	19	4	12.27	1.25	.46	-.82	.36	-.70	O .94	.39	100.0	76.7	S5
7	19	4	12.27	1.25	.46	-.82	.36	-.70	P .94	.39	100.0	76.7	S7
15	19	4	12.27	1.25	.46	-.82	.36	-.70	Q .94	.39	100.0	76.7	S15
16	11	4	-7.81	1.23	.44	-.83	.34	-.73	R .94	.39	100.0	76.4	S16
19	11	4	-7.81	1.23	.44	-.83	.34	-.73	S .94	.39	100.0	76.4	S19
23	11	4	-7.81	1.23	.44	-.83	.34	-.73	T .94	.39	100.0	76.4	S23
25	11	4	-7.81	1.23	.44	-.83	.34	-.73	t .94	.39	100.0	76.4	S25
28	11	4	-7.81	1.23	.44	-.83	.34	-.73	s .94	.39	100.0	76.4	S28
2	12	4	-5.45	1.87	.04	-.90	.04	-.95	r .00	.21	100.0	92.5	S2
3	12	4	-5.45	1.87	.04	-.90	.04	-.95	q .00	.21	100.0	92.5	S3
4	12	4	-5.45	1.87	.04	-.90	.04	-.95	p .00	.21	100.0	92.5	S4
13	12	4	-5.45	1.87	.04	-.90	.04	-.95	o .00	.21	100.0	92.5	S13
17	12	4	-5.45	1.87	.04	-.90	.04	-.95	n .00	.21	100.0	92.5	S17
29	12	4	-5.45	1.87	.04	-.90	.04	-.95	m .00	.21	100.0	92.5	S29
8	16	4	4.59	7.63	.00	-1.36	.00	-1.38	l .00	.05	100.0	99.6	S8
9	16	4	4.59	7.63	.00	-1.36	.00	-1.38	k .00	.05	100.0	99.6	S9
11	16	4	4.59	7.63	.00	-1.36	.00	-1.38	j .00	.05	100.0	99.6	S11
12	16	4	4.59	7.63	.00	-1.36	.00	-1.38	i .00	.05	100.0	99.6	S12
14	16	4	4.59	7.63	.00	-1.36	.00	-1.38	h .00	.05	100.0	99.6	S14
31	16	4	4.59	7.63	.00	-1.36	.00	-1.38	g .00	.05	100.0	99.6	S31
32	16	4	4.59	7.63	.00	-1.36	.00	-1.38	f .00	.05	100.0	99.6	S32
33	16	4	4.59	7.63	.00	-1.36	.00	-1.38	e .00	.05	100.0	99.6	S33
34	16	4	4.59	7.63	.00	-1.36	.00	-1.38	d .00	.05	100.0	99.6	S34
36	16	4	4.59	7.63	.00	-1.36	.00	-1.38	c .00	.05	100.0	99.6	S36
38	16	4	4.59	7.63	.00	-1.36	.00	-1.38	b .00	.05	100.0	99.6	S38
39	16	4	4.59	7.63	.00	-1.36	.00	-1.38	a .00	.05	100.0	99.6	S39
MEAN	14.2	4.0	.04	3.22	.59	-.5	.57	-.5			84.4	83.7	
P.SD	2.7	.0	6.87	2.90	.76	1.1	.79	1.0			24.8	13.0	

FIGURE 2. Person Statistics: Misfit Order

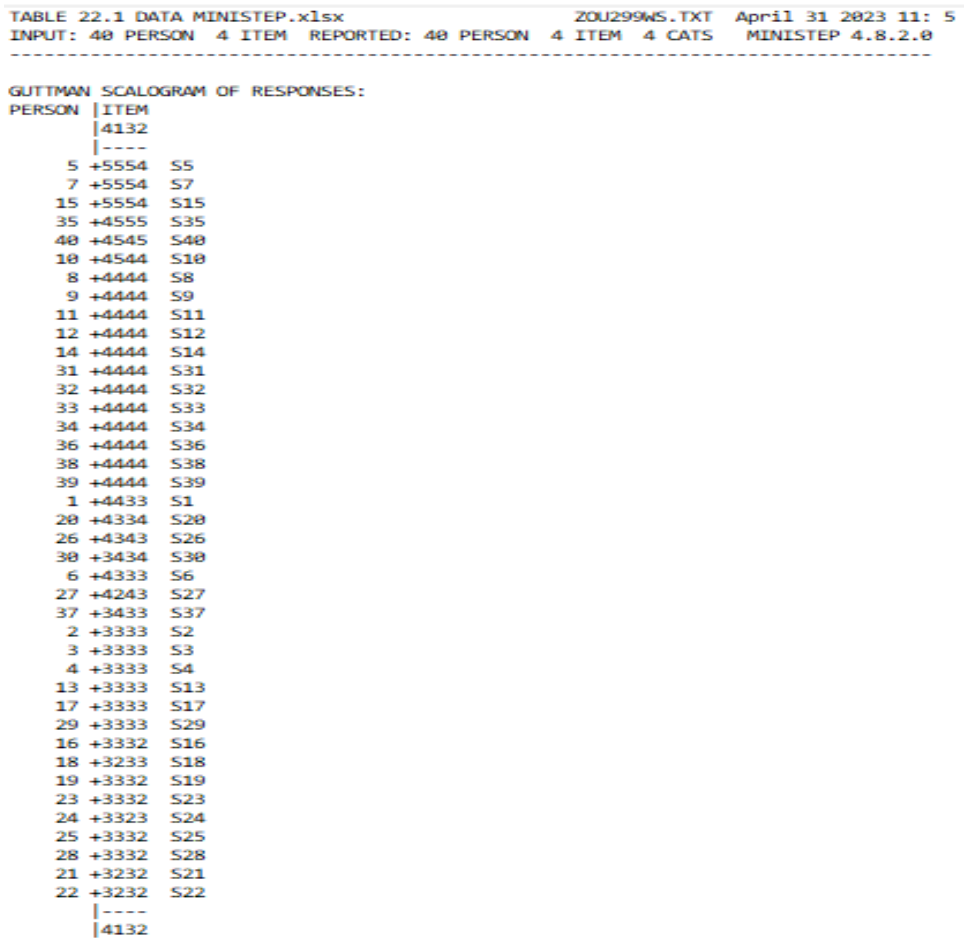
According to Linacre (2002), there are two statistics that can be used to assess the fit of data to the Rasch model, namely infit (inlier-sensitive or information-weighted fit) and outfit (outlier-sensitive or information-weighted fit). These statistics are commonly reported in the form of mean squares (MNSQ) and standardized z-values (ZSTD). MNSQ represents the average of squared residuals for an item, while ZSTD (standardized form) is a transformation of the average squared values with sample size correction (Bond & Fox, 2015). Therefore, in this study, to determine whether an item or respondent (person) fits or misfits the Rasch model, the output of the Winsteps Rasch software, specifically the Outfit Mean Square (MNSQ) statistics, needs to be interpreted. MNSQ statistics are chosen because they are independent of sample size.

Table 3. Person Misfit Summary

Students	Outfit MNSQ	Students	Outfit MNSQ
27	3.06	4	0.4
45	2.64	13	0.4
40	2.18	17	0.4
30	2.16	29	0.4
20	1.75	8	0.0
5	0.36	9	0.0
7	0.36	11	0.0
15	0.36	12	0.0
16	0.34	14	0.0
19	0.34	31	0.0
23	0.34	32	0.0
25	0.34	33	0.0
28	0.34	34	0.0
2	0.4	36	0.0
3	0.4	38	0.0
39	0.0		

Linacre (2002) provides a rule of thumb to assess the implications of model fit on measurement. An MNSQ value greater than 2.0 indicates a damaging effect on the measurement

system, while a value between 1.5 and 2.0 suggests a lack of meaning in the measurement. MNSQ values between 0.5 and 1.5 are considered beneficial for measurement, and an MNSQ value below 0.5 is not useful for measurement but does not disrupt the measurement system. Figure 2 presents a summary of difficulty levels and outfit MNSQ statistics. According to Linacre's criteria (2002), person fit statistics are interpreted. The results show that 31 out of 41 students are classified as person misfits, namely students with identification numbers 27, 35, 40, 30, 20, 15, 16, 19, 23, 25, 28, 2, 3, 39, 4, 13, 17, 29, 8, 9, 11, 12, 14, 31, 32, 33, 34, 36, and 38. This means that the abilities of these 31 students have response patterns that cannot be predicted by the model (Smith, 2001). However, response patterns can depict the accuracy of each student's response to each item (Sumintono & Widhiarso, 2015). One way to identify the causes of person misfit is through Guttman matrices or scalograms. Guttman matrices provide valuable information as the items have been ordered from the easiest item, mechanic (4), to diction (3), structure (2), and content (1). These matrices can also indicate the unidimensionality of the data (Hambleton & Swaminathan, 1991). Below is the scalogram of the 40 students based on the identification of difficulty levels from lowest to highest. Identification of students classified as person misfits based on the Guttman matrices.



```
TABLE 22.1 DATA MINISTEP.xlsx                    ZOU299WS.TXT  April 31 2023 11: 5
INPUT: 40 PERSON  4 ITEM  REPORTED: 40 PERSON  4 ITEM  4 CATS   MINISTEP 4.8.2.0
-------------------------------------------------------------------------------

GUTTMAN SCALOGRAM OF RESPONSES:
PERSON |ITEM
       |4132
       |----
     5 +5554  S5
     7 +5554  S7
    15 +5554  S15
    35 +4555  S35
    40 +4545  S40
    10 +4544  S10
     8 +4444  S8
     9 +4444  S9
    11 +4444  S11
    12 +4444  S12
    14 +4444  S14
    31 +4444  S31
    32 +4444  S32
    33 +4444  S33
    34 +4444  S34
    36 +4444  S36
    38 +4444  S38
    39 +4444  S39
     1 +4433  S1
    20 +4334  S20
    26 +4343  S26
    30 +3434  S30
     6 +4333  S6
    27 +4243  S27
    37 +3433  S37
     2 +3333  S2
     3 +3333  S3
     4 +3333  S4
    13 +3333  S13
    17 +3333  S17
    29 +3333  S29
    16 +3332  S16
    18 +3233  S18
    19 +3332  S19
    23 +3332  S23
    24 +3323  S24
    25 +3332  S25
    28 +3332  S28
    21 +3232  S21
    22 +3232  S22
       |----
       |4132
```

FIGURE 3. Guttman Scalogram of Responses

Based on the presentation of the Guttman matrix above, it can be concluded that students with the identification numbers 35, 40, 10, 30, and 37 are considered person misfits in the Rasch model. This is because these students exhibit unusual response patterns, achieving high scores on more difficult items such as content and structure (Jacobs et al, 1981), but obtaining low scores on easier items like diction and mechanics. According to the definition of the Rasch model, which states that students with lower abilities should not have a high likelihood of correctly answering more challenging items, it can be inferred that there may have been an error in assessing the students' writing skills. This identification result aligns with Meijer (1996) and Karabatsos (2003), who mention at least five possible causes of person misfit. These include cheating (such as copying answers from other test-takers), where unfair behaviour leads to correct answers on items that the student couldn't have answered correctly; careless responding, which occurs when test-takers answer difficult items correctly but answer easy items incorrectly in an unclear manner; lucky guessing, when test-takers randomly guess the correct answer on items they don't actually know; creative responding, which only occurs among high-ability test-takers who respond incorrectly to easy items because they interpret the items in a

unique and creative way; and finally, random responding, which refers to situations where test-takers randomly choose multiple-choice options when responding to items. Furthermore, the Rasch model identified three students with identification numbers 5, 57, and 15 as individuals who did not fit well within the model. This determination was made due to their extreme scores, which led to unmeasurable fit statistics, indicating an overfit. According to Meijer (1996), measuring person fit not only identifies impossible response patterns but also patterns that are too likely. The Rasch model predicts uncertainty, and having too much certainty actually indicates limitations in responses.

D. CLOSING

Based on the analysis results using the Outfit MNSQ range, it was found that there were 31 students who were classified as person misfit. This indicates that there is a discrepancy in the understanding and application of the 4 skills in academic writing taught to the students. As a result, these 31 students are unable to effectively apply the sub-parameters or items (Jacob et al) of the four parameters in academic writing. Additionally, anomalies were found in the students' response patterns to more difficult items such as content writing and correct structure application. Both of these items received significantly higher scores compared to the other two items, mechanics and diction, which are relatively easier for the students to master. There are 5 students who fall into this misfit category, as shown in the scale map by better scores on the more difficult items

E. REFERENCE

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(1), 561–573. https://doi.org/https://doi.org/10.1007/BF02293814

Bond, T., & Fox, C. (2015). *Applying the Rasch Model*. Routledge. https://doi.org/https://doi.org/10.4324/9781315814698

Bond, T.G., & Fox, C. M. (2007). *pplying the Rasch model: Fundamental measurement in the human sciences (2nd ed.)*. awrence Erlbaum Associate.

Boone, W. J., Staver, J. R., Yale, M. S., Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Item Measures. Rasch Analysis in the Human Sciences*. 93–110. https://doi.org/https://doi.org/10.1007/978-94-007-6857-4_5

Engelhard, G. (1992). The measurement of writing ability with a many-faceted Rasch model. *Applied Measurement in Education*, *5*(3), 171–191.

Fisher, W. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *1*.

Hambleton, R.K., Swaminathan, H., & R. H. J. (1991). *Fundamental of item response theory*. Sage Publishing.

Jacobs., Holly. L., Stephen, A., Zingkgraf., Deanne. R., Wormuth, V., Faye, H., Jane, B., H. (1981). *Testing ESL Composition: A Practical Approach*. Newbury House Publishers, Inc.

Karabatsos, G. (2003). Comparing the abberant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, *16*(4), 277–298.

Li, G., Pan, Y., & Wang, W. (2021). *Using*

Generalizability Theory and Many-Facet Rasch Model to Evaluate In-Basket Tests for Managerial Positions. *12*(July), 1–10. https://doi.org/10.3389/fpsyg.2021.660553

Linacre, J. M. (2002). KR-20/Cronbach alpha or Rasch person reliability: Which tells us the truth? *Rasch Measurement Transactions*, *11*, 580–581.

Linacre, J. M. (. (2002). What do infit and outfit mean-square and standardized mean? *Rasch Measurement Transaction*, *16*, 878.

Meijer, R. R. (1996). Person-fit research: an introduction. *Applied Measurement in Education*, *9*(1), 3–8.

Misbach, I. H., & Sumintono, B. (2014). Pengembangan dan Validasi Instrumen "Persepsi Siswa Tehadap Karakter Moral Guru" di Indonesia dengan Model Rasch. . . *PROCEEDING Seminar Nasional Psikometri*, 148–162.

Osman, S. A., Naam, S. I., Jaafar, O., Badaruzzaman, W. H. W., & Rahmat, R. A. A. O. K. (2012). Application of Rasch Model in Measuring Students' Performance in Civil Engineering Design II Course. *Procedia - Social and Behavioral Sciences*, *56*, 59–66. https://doi.org/10.1016/J.SBSPRO.2012.09.632

Rahman, Y. A. (2023). *Person and Item Validity and Reliability in Essay Writing Using Rasch Model*. *15*(1), 41–55. https://doi.org/10.35457/konstruk.v15i1.2618

Razak, N. bin Abd, Khairani, A.Z. bin, & Thien, L. M. (2012). Examining quality of mathemtics test items using rasch model: Preminarily analysis. *Procedia - Social and Behavioral Sciences*, *69*, 2205–2214. https://doi.org/https://doi.org/10.1016/j.sbspro.2012.12.187

Sumintono, B. & Widhiarso, W. (2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Trim Komunikata Publishing House.

Tan, S. (2013). Validation of an Analytic Rating Scale for Writing: A Rasch Modeling Approach. *Tabaran Institute of Higher Education. Iranian Journal of Language Testing*, *3*(1).

Wibisono, S. (2018). Aplikasi Model Rasch Untuk Validasi Instrumen Pengukuran Fundamentalisme Agama Bagi Responden Muslim. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, *5*(1). https://doi.org/https://doi.org/10.15408/jp3i.v5i1.9239