Investigating The Central Tendency Effect Among Raters in Complex Problem-Solving Skills in Physics Essay Items: An Application of Many-Facet Rasch Measurement

Ekwere, Ndifreke Soni
Department of Educational Evaluation and Counselling Psychology, University of Benin, Edo State, Nigeria.
ndifreke.soni@educ.uniben.edu

Adeosun, Praise Kehinde (PhD)
Department of Educational Evaluation and Counseling Psychology
University of Benin, Edo State, Nigeria.
E-mail: praise.adeosun@uniben.edu
GSM: +234-8060481557
https://orcid.org/0009-0003-1870-978X

Google Scholar Account: https://scholar.google.com/citations?hl=en&users=uRLbDpYAAAAJ

ABSTRACT

The study sought to investigate the central tendency effect among raters in complex problem-solving skills in physics essay items using Many-Facet Rasch Measurement. This became necessary because the central tendency effect can affect students' scores.

The research design adopted for this study is a descriptive research design based on the survey method. The population of the study comprised eighty-eight (88) public schools in all the local government areas with a physics student population of 3642 students and ninety-four (94) physics teachers in all the Senior Secondary Schools in Uyo Senatorial District for the 2022/2023 academic session. 364 SS3 physics students and 37 physics teachers from the 31 selected secondary schools in Uyo Senatorial District were sampled using a multistage sampling technique for effective selection. The multi-stage sampling technique was adopted for the study. The instrument used for data collection was the Physics Achievement Test (PAT) obtained from WAEC and NECO 2020 Physics Essay items. However, the reliability coefficient of 0.91 and 0.90 respectively.

The finding revealed that there was no central tendency effect error committed by individual raters when rating complex problem-solving skills in WAEC and NECO 2020 physics Essay Items. Also, reveals there is a significant difference at the rater's group level exhibiting central tendency effect when rating complex...
problem-solving skills between WAEC and NECO 2020 physics Essay Items which implies that there is no group-level central tendency effect present among raters when rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items. The researchers concluded that rater effects are sources of variance in performance ratings that are associated with the rater's behaviour and not the actual performance of the ratee. It was recommended that to prevent the central tendency effect, it is important for superior raters to regularly observe their subordinates throughout the evaluation period and keep notes for future reference.

**Keywords**: Central Tendency effect, Raters, Complex problem solving skills, Many Facet-Rasch Measurement

**INTRODUCTION**

The National Examination Council (NECO) is an examination body established in the year 1999 in Nigeria. The National Examination Council engaged in conducting the Senior Secondary Certificate Examination (SSCE). The Senior Secondary Certificate Examination is for students who are at the apex of the Secondary School level. NECO is responsible for organizing and administering various examinations, including the Senior School Certificate Examination (SSCE) for both internal and external candidates. This examination assesses the performance of secondary school students in Nigeria. NECO develops and formulates examination questions for the subjects offered in the SSCE. These questions are designed to evaluate students' understanding and knowledge in their respective fields of study. NECO is responsible for issuing certificates to candidates who successfully pass the SSCE. These certificates are recognized by educational institutions, employers, and government agencies in Nigeria.

In the same vein, the West African Examinations Council (WAEC) is a prominent educational organization responsible for conducting and coordinating standardized examinations in English-speaking West African countries. Founded in 1952, WAEC has played a significant role in shaping the education systems of its member countries and ensuring the integrity of assessment processes. WAEC is responsible for the administration of various examinations, including the West African Senior School Certificate Examination (WASSCE) for both school and private candidates. This examination assesses the performance of secondary school students in West Africa. WAEC develops and sets examination questions for the various subjects offered in the WASSCE. These questions are designed to evaluate students' understanding and knowledge in their respective fields of study. After the examinations are conducted, WAEC is responsible for the meticulous process of marking and grading the answer scripts to ensure an accurate and fair assessment of students' performance.

Achievement tests are administered to the students to ascertain their strengths and abilities. A kind of assessment known as an achievement test measures a person's knowledge, abilities, or comprehension in a particular field or subject. It attempts to ascertain what a person has learnt or accomplished after receiving
Investigating The Central Tendency Effect Among Raters in Complex Problem-Solving Skills in Physics Essay Items: An Application of Many-Facet Rasch Measurement

instruction or learning. To assess a student's knowledge of a particular curriculum or set of learning objectives, achievement tests are usually used in educational settings. These tests give teachers information about how well a student has understood the topic being taught and can aid in the development of curricula and instruction. To ensure fairness and uniformity in evaluation, achievement tests are given in a standardized way.

To evaluate a person's knowledge, skills, or competencies in a particular field, like physics, an achievement test assessment is involved thereby evaluating and analyzing the test results. This procedure tries to assess a test-taker's level of mastery concerning particular learning goals or academic criteria. Eliminating any possible errors in the rating process is necessary to develop appropriate assessment practices.

Complex problem-solving skills are considered a primary goal in physics instruction for both secondary and university physics courses. Problem-solving is the process of moving toward a goal when the path to the goal is not known or uncertain. Therefore, complex problem-solving skills are those steps or strategies taken to aid and attain a conceived aim or goal in solving problems. Complex problem-solving also depends on a student's prior experience, as some skills can be learned well enough that they become automatic and require minimal effort. According to Wendorf (2019), problem-solving skills include identifying the problem (identifying the nature of the problem by visualizing the physical situation, and translating the written information in the problem into mathematical variables or drawing a diagram showing the objects, and their motions or interactions), define the main elements of the problem (on the diagram, label all the known and unknown information), examine possible solutions and act on resolving the problem (this often includes working through previous similar problems and observing the solution process).

Many-Facet Rasch Measurement (MFRM) models, also known as multifaceted models, incorporate more variables or aspects (such as assessors, scoring criteria, and tasks) than two variables or aspects the edge introduced into the classical test situation is the candidate and the element. One of the key features of the MFRM model is its ability to evaluate the quality of measurements in multiple dimensions. This is done through the use of multiple aspects, which are defined as different aspects of the test procedure that can affect the quality of the measurement. Facets are different sources of variation that can affect the measurement process.

MFRM is a model-based psychoanalysis that has many applications in performance evaluation. It is an extension of Rasch's basic single-parameter item response theory model, a type of psychometric model used to estimate students' ability and test item difficulty on the same scale. In MFRM, all Facets are placed on the same logit scale, allowing comparisons between aspects. Secondly, the MFRM model provides projected model-driven estimates of the grades students should have received, after accounting for measurement errors related to all included dimensions.

The MFRM is designed to account for multiple sources of variability, including rater severity, examinee ability, and item difficulty. This multidimensional approach allows for a more comprehensive
understanding of the data, which is crucial when dealing with complex assessments. The MFRM provides facet-specific information about each rater's behaviour, each examinee's ability, and each item's difficulty. This granularity enables a detailed analysis of how central tendency effects may be manifesting across different facets of the assessment. The MFRM separates the facets (e.g., raters, items, examinees) in the analysis, allowing for a clear examination of their contributions to the measurement process. This separation helps in identifying and quantifying central tendency effects specifically related to raters.

Rater's error is a systematic error that occurs when teachers assess and evaluate students. Personal perceptions and biases can affect how we evaluate an individual's performance. Over the last century, examining these rating patterns within and across raters has received tremendous research focus to determine the degree to which the ratings can be accurate indicators of the examinees' abilities (Foltz, & Rosenstein, 2018). Myford and Wolfe (2003) categorized five major types of rater effects: leniency/severity, randomness, halo, and central tendency.

Central tendency is otherwise known as restriction of range, which can present itself in different ways. In some cases, the rater who tends to overuse the middle categories of the scale may be able to accurately assess the level of performance of the very highest and lowest performing rates that is those whose performance measures fall at the extreme upper and lower ends of the performance continuum. Central tendency occurs when the rater awards an average score to an examinee that may have scored higher or lower. For example, when the rater (teacher) awards an examinee with a score of 60% the examinee (student) should have scored 80% or 40%.

Advancement of the present study to the previous study

The present study is the advancement from the previous publication on "Application of Many-Facet Rasch Measurement in complex problem-solving skills in 2019 NECO physics essay items" in the sense that this present paper specifically targets the tendency of raters to cluster their ratings around the midpoint of the scale. This addresses a different type of bias where raters may be inclined to assign scores that are closer to the average, regardless of the true performance of the examinees. This will enable educational institutions and testing agencies to implement measures to mitigate central tendency effects and enhance the fairness and accuracy of assessments.

Statement of the Problem

When assessing student performance in solving complex problems, there are student and environmental factors, as well as unforeseen characteristics that influence results. These unexpected features are collectively called scoring errors. Marker errors are caused by inaccuracies on the part of teachers or examiners and affect the reliability and validity of marking.

Raters’ errors can diminish the quality of feedback provided to students. When physics raters are not accurate in identifying and assessing the strengths and weaknesses in students' complex problem-solving...
approaches, the feedback given may not be helpful or constructive. This can impede students' learning and
growth in complex problem-solving skills. Rater errors, particularly if they consistently undervalue or
underestimate students' problem-solving skills, can negatively impact student motivation. Students who
consistently receive lower ratings than they deserve may become demotivated and disengaged, leading to a
decline in their efforts to develop and improve their problem-solving abilities.

The error-prone nature of most aspects of the measurement especially that of human raters, raises
serious concerns about the psychometric quality of the marks given to test-takers. These concerns should be
carefully considered, especially for high-risk examinations such as her WAEC and NECO, as the results can
have a significant impact on candidates' careers and study plans. With this in mind, therefore, we apply many
facets of Rasch measurement models to detect central tendency effects in assessing complex problem-solving
skills in WAEC and NECO physics essay items.

**Purpose of the Study**

The purpose of this study is to detect central tendency effect errors among raters in complex problem-

**Research Questions**

First, to what extent does the individual rater produce a central tendency effect when rating complex
problem-solving skills in WAEC and NECO 2020 physics Essay Items?

Second, is there a difference in the central tendency effect at the rater's group level when rating
complex problem-solving skills in WAEC and NECO 2020 physics Essay Items?

**Hypothesis**

There is no significant difference at the rater's group level exhibiting a central tendency effect when
rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items.

**Significance of the Study**

The results of the research may be useful to students as they reduce manipulation of results caused by
rater error. The finding could be beneficial to test constructors as students' performance could be enhanced
positively by detecting multiple errors caused by the rater which will increase the reliability of the scores.

**METHODOLOGY**

**Research Design**

The research design adopted for this study is a descriptive research design based on the survey method.
Investigating The Central Tendency Effect Among Raters in Complex Problem-Solving Skills in Physics Essay Items: An Application of Many-Facet Rasch Measurement

Target Population
The population of the study comprised eighty-eight (88) public schools in all the local government areas with a physics student population of 3642 students and ninety-four (94) physics teachers in all the Senior Secondary Schools in Uyo Senatorial District for the 2022/2023 academic session.

Sample and Sampling Techniques
A sample of three hundred and sixty-four (364) Senior Secondary three (SS3) physics students and thirty-seven (37) physics teachers from the thirty-one (31) selected secondary schools in Uyo Senatorial District was used for the study. The multi-stage sampling technique was adopted for the study.

Firstly, proportionate sampling was used to select 31 schools representing 35% of the total number of schools in each of the local government areas used in this study.

Secondly, simple random sampling was used to select 364 physics students representing 10% of the physics students from the 31 schools in the Uyo senatorial district.

Thirdly, a stratified sampling technique was used to obtain the sample size of 37 physics teachers in 31 schools. This was done by dividing the population into strata based on gender. Then randomly sample 40% of each of the strata.

Data Collection Instrument
The instrument used for data collection was the Physics Achievement Test (PAT) obtained from WAEC and NECO 2020 Physics Essay items.

Validity and Reliability of the Instrument
The easy items were validated by the Examination Development Department of WAEC and NECO. The reliability of the instruments was established using a sample of 30 students from public Senior Secondary (SS 3) who were not part of the sample but of the main study population. The reliability of NECO and WAEC was determined using fit statistics in Winsteps version 4.8.2.0 package to obtain a reliability coefficient of 0.91 and 0.90 respectively.

Data Collection
The researcher administered the instrument (Physics Achievement Test) to SSS 3 students with the assistance of their physics teachers in the selected schools. The duration for the test was 1 hour 30 minutes. After administering the test, the responses were retrieved from the students. The students' responses were rated by the 37 physics teachers. A 6-point scale was used. 0-39 'fail' (1), 40 – 44 'fair' (2), 45 – 49 'pass' (3), 50-59 'good' (4), 60 – 69 'very good (5) and 70 and above 'excellent' (6). After scoring and rating the test by the
Investigating The Central Tendency Effect Among Raters in Complex Problem-Solving Skills in Physics Essay Items: An Application of Many-Facet Rasch Measurement

physics teachers, the researcher retrieves all the student’s responses from the raters (physics teachers) for proper analysis.

Data Analysis

The data collected were analyzed using Winsteps version 5.1.1.0 software for FACET. Research question 1 was answered using descriptive statistics while hypothesis 1 was tested using fixed Chi-Square at the alpha level of 0.05.

RESULTS

Research Question 1: To what extent does the individual rater produce a central tendency effect when rating complex problem-solving skills in WAEC and NECO 2020 physics Essay Items?

Table 1: Category Statistics that Show the Extent to which Individual Raters Produce Central Tendency Effect when Rating Complex Problem Solving Skills in WAEC and NECO 2020 Physics Essay Items.

<table>
<thead>
<tr>
<th>Rater (R)</th>
<th>Measure(logits)</th>
<th>Standard Error</th>
<th>Infit Mean Square (MNSQ)</th>
<th>Outfit Mean Square (MNSQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.27</td>
<td>0.54</td>
<td>1.21</td>
<td>1.14</td>
</tr>
<tr>
<td>2</td>
<td>-3.07</td>
<td>0.55</td>
<td>2.44</td>
<td>2.33</td>
</tr>
<tr>
<td>3</td>
<td>-3.68</td>
<td>0.55</td>
<td>1.36</td>
<td>4.68</td>
</tr>
<tr>
<td>4</td>
<td>-0.99</td>
<td>0.53</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>5</td>
<td>-0.72</td>
<td>0.52</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>6</td>
<td>-1.86</td>
<td>0.55</td>
<td>1.02</td>
<td>0.98</td>
</tr>
<tr>
<td>7</td>
<td>-4.68</td>
<td>0.58</td>
<td>0.81</td>
<td>0.67</td>
</tr>
<tr>
<td>8</td>
<td>-1.27</td>
<td>0.55</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>9</td>
<td>-5.85</td>
<td>0.71</td>
<td>0.59</td>
<td>0.38</td>
</tr>
<tr>
<td>10</td>
<td>-0.72</td>
<td>0.52</td>
<td>0.38</td>
<td>0.34</td>
</tr>
<tr>
<td>11</td>
<td>-1.27</td>
<td>0.54</td>
<td>1.21</td>
<td>1.14</td>
</tr>
<tr>
<td>12</td>
<td>-3.07</td>
<td>0.55</td>
<td>2.50</td>
<td>2.33</td>
</tr>
<tr>
<td>13</td>
<td>-3.68</td>
<td>0.55</td>
<td>1.36</td>
<td>4.68</td>
</tr>
<tr>
<td>14</td>
<td>-0.99</td>
<td>0.53</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>15</td>
<td>-0.72</td>
<td>0.52</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>16</td>
<td>-1.86</td>
<td>0.55</td>
<td>1.02</td>
<td>0.98</td>
</tr>
<tr>
<td>17</td>
<td>-4.68</td>
<td>0.58</td>
<td>0.81</td>
<td>0.67</td>
</tr>
<tr>
<td>18</td>
<td>-1.27</td>
<td>0.55</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>19</td>
<td>-5.85</td>
<td>0.71</td>
<td>0.59</td>
<td>0.38</td>
</tr>
<tr>
<td>20</td>
<td>-0.72</td>
<td>0.52</td>
<td>0.38</td>
<td>0.34</td>
</tr>
<tr>
<td>21</td>
<td>-1.27</td>
<td>0.54</td>
<td>1.21</td>
<td>1.14</td>
</tr>
</tbody>
</table>
In detecting the central tendency effect, we consider the Infit Mean Square and Outfit Mean Square. When the Infit and Outfit Mean Square is between 0.5 – 1.5 indicate no central effect. But if the fit statistics is greater than 1.5 or less than 0.5, this indicates the presence of a central tendency effect. Table 1 shows 15 raters: rater 1, 4, 6, 7, 11, 14, 16, 17, 21, 24, 26, 27, 31, 36, and 37 with the Infit Mean Square and Outfit Mean Square of 1.21 & 1.14, 0.85 & 0.84, 1.02 & 0.98, 0.81 & 0.67, 1.21 & 1.14, 0.85 & 0.84, 1.02 & 0.98, 0.81 & 0.67, 1.21 & 1.14, 0.85 & 0.84, 1.02 & 0.98, 0.81 & 0.67, 0.81 & 0.67 respectively. These values fit into the model indicating that these raters did not commit a central tendency effect when rating complex problem-solving skills in WAEC and NECO 2020 physics Essay Items.

**Hypothesis one**: There is no significant difference at the rater's group level exhibiting a central tendency effect when rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items.

**Table 2: Group – Level indices of Central Tendency Effect in Complex Problem Solving Skills Between WAEC and NECO 2020 Physics Essay Items.**

<table>
<thead>
<tr>
<th>Indices</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratee Separation Ratio</td>
<td>4.35</td>
</tr>
<tr>
<td>Ratee Separation Index</td>
<td>6.18</td>
</tr>
<tr>
<td>Ratee Separation Reliability</td>
<td>0.95</td>
</tr>
<tr>
<td>Fixed-effect Chi-Square Statistics</td>
<td>217 (df=13, P&lt;0.05), Sig. =0.00</td>
</tr>
</tbody>
</table>
Table 2 shows a Chi-square value of 217 and a P-value of 0.00. Testing at an alpha level of .05, the P-value is less than the alpha level, so the null hypothesis which states that there is no significant difference at the rater's group level exhibiting central tendency effect when rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items is rejected. Therefore there is a significant difference at the rater's group level exhibiting a central tendency effect when rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items. This implies that there is no group-level central tendency effect present among raters when rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items.

Also, Table 2 shows the Ratee Separation Ratio of 4.35. This implies that the indicator did not suggest a group-level central tendency effect among raters when rating student scores in complex problem-solving skills between WAEC and NECO 2020 physics Essay Items. Table 2 also shows the Ratee Separation Index of 6.18. This indicator does not suggest a group level of central tendency among raters when rating student scores in complex problem-solving skills between WAEC and NECO 2020 physics Essay Items. Furthermore, Table 2 reveals a Ratee Separation Reliability of .95. The high degree of Ratee separation reliability indicates that there was no group-level central tendency among raters when rating student scores in complex problem-solving skills between WAEC and NECO 2020 physics Essay Items.

DISCUSSION OF FINDINGS

The finding from research question one reveals that there was no central tendency effect error committed by individual raters when rating complex problem-solving skills in WAEC and NECO 2020 physics Essay Items. This may be due to the lack of familiarity with the skills being evaluated, and a desire to avoid extreme ratings. The absence of a central tendency effect may be because clear criteria were provided to assess the quality of the response. These criteria help raters evaluate the response more objectively and reduce the likelihood of central tendency effect when rating complex problem-solving skills in WAEC and NECO 2020 physics Essay Items. Also in complex problem-solving tasks, there were wide ranges of responses that are not easily comparable. This variability makes it difficult for raters to commit the central tendency effect.

The finding is not in support of the findings of Rahman, et al. (2017) who used Many Facet Rasch Measurement (MFRM) to explore rater error focusing on the central tendency effect. The results at the individual level indicated that two raters exhibit centrality. Furthermore, the findings do not coincide with the study of Esfandiari (2015), The result showed that individual peer assessors committed varying degrees of central tendency effect. The finding was against the study of Leckie and Baird (2011) who examined rater effects on essay scoring in an operational monitoring system from England's 2008 national curriculum English writing test for 14-year-olds. The result revealed that there was a central tendency effect. However, the finding aligns with the study of Farrokhi et al. (2011) who apply the many-facet Rasch model (MFRM) to detect the central tendency effect in self-assessment, peer assessment and teacher assessment. The results showed that the three types of assessor-self-assessor, peer-assessor and teacher-assessor did not exhibit any sign of centrality at the individual level.
The findings from hypothesis one reveal that there is a significant difference at the rater's group level exhibiting a central tendency effect when rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items. This implies that there is no group-level central tendency effect present among raters when rating complex problem-solving skills between WAEC and NECO 2020 physics Essay Items. This may be because of the rater's sense of accountability and professionalism in their role, recognizing the importance of providing accurate and fair evaluations. This mindset can help reduce the central tendency effect, as raters strive to assign scores that accurately reflect the performance of the students, rather than defaulting to the middle of the rating scale. The Ratee Separation Ratio indicates that the spread of ratee performance measures is over 4 times larger than the precision of those measures. The Ratee Separation Index connotes that there are about 6 statistically distinct strata of ratee performance in the sample of ratees. The Ratee Separation Reliability implies that the ratees are well differentiated in terms of their levels of performance.

The finding was in line with the study of Rahman et al. (2017) who employed Many Facet Rasch Measurement (MFRM) to explore rater error focusing on the central tendency effect. The result reveals that at the group level, none of the separation statistics indicated centrality. However, the finding was similar to the findings of Esfandiari (2015) who applied the Many-Facet Rasch Measurement (MFRM) to detect two pervasive rater errors among peer-assessor's rating EFL essays. The result reveals that the ratings of peer assessors and those of teacher-assessors were not statistically significantly different. The finding corresponds with the study of Farrokhhi et al. (2011) who employed the Many-Facet Rasch model (MFRM), to detect central tendency effects in self-assessment, peer assessment and teacher assessment. The results showed that the three types of assessor-self-assessor, peer-assessor and teacher-assessor did not exhibit any sign of centrality either at the group level.

Conclusion

Based on the findings, it was concluded that at the individual level, raters did not commit the central tendency effect when rating complex problem-solving skills in WAEC and NECO 2020 physics Essay Items. Also, at the group-level rating, raters did not commit a central tendency effect when rating complex problem-solving skills in WAEC and NECO 2020 physics Essay Items. It can be deduced that rater effects are sources of variance in performance ratings that are associated with the rater's behaviour and not the actual performance of the ratee.

Recommendation

To prevent the central tendency effect, it is important for superior raters to regularly observe their subordinates throughout the evaluation period and keep notes for future reference.
REFERENCES


