

Classroom Teachers' Bias in Assessing Disadvantaged Primary School Students' Reading Comprehension Skills

Yusuf Kızıllaş¹, Mehmet Şata^{✉2} & Fuat Elkonca³

¹ Department of Primary School Education, Van Yuzuncu Yil University, Van, Turkey

² Department of Measurement and Assessment, Agri Ibrahim Cecen University, Ağrı, Turkey

³ Department of Measurement and Assessment, Mus Alparslan University, Muş, Turkey

✉ mehmetwsata@gmail.com

Abstract: It is known that the reading performance of disadvantaged students is lower when compared to non-disadvantaged students. It has always been discussed that being disadvantaged affects teachers' bias in scoring students' reading performance. Therefore, the existence and effect of the teacher factor in the low level of reading performance of students in disadvantaged groups is worth investigating. The reading comprehension skills of disadvantaged and non-disadvantaged fourth grade primary school students were assessed in this study. Moreover, whether the classroom teachers were biased in scoring the reading comprehension skills of the disadvantaged students were also investigated. The reading comprehension skills of 48 students, 13 of whom were disadvantaged and 35 were non-disadvantaged, were assessed by three classroom teachers (raters) using an analytical rubric. The data were analyzed using the Many-facet Rasch measurement model. It was found that the classroom teachers were biased in evaluating the reading comprehension skills of disadvantaged students. Likewise, the reading comprehension levels of the non-disadvantaged students were higher than the disadvantaged students.

Keywords: Disadvantaged primary school students, Reading comprehension, Reading Assessment, Rater bias, Many-facet Rasch.

1. Introduction

Reading is known as one of the significant language skills. The dynamics of the reading skill makes it unique. Akyol's (2019) arguments support the significance of the reading skill. In the context of these explanations, reading is considered as a meaning-making process, which requires the use of prior knowledge, includes a dynamic interaction between the reader and the author, and is carried out in an organized environment. What is meant by meaning-making is to establish a link between the text and the reader's prior knowledge. This link has an important role in both initiating and accelerating the reading comprehension process. Reading comprehension (Solari et al., 2018) skill, which includes understanding, comprehending, and making sense of a text, is the main purpose of reading (Grabe & Stoller, 2002). In this respect, it is also important to discuss the process of reading comprehension and the evaluation of this process. In this respect, it is also important to discuss the process of reading comprehension and its assessment. So much so that reading comprehension is considered the most important skill acquired in primary school (Baddour, 2019).

Therefore, it is necessary to explain reading comprehension not as a simple act such as decoding the codes in the text, but as constructing meaning through the text. In other words, reading comprehension is a process in which the messages and meanings in the text are reconstructed by the reader in the context of various strategies and skills (Habib, 2016). According to the Report of the National Reading Panel (2000), which explains the significance of reading comprehension, teachers' role in reading comprehension needs to be thoroughly explored because they prepare students to the reading comprehension process by getting involved in the assessment and management of this process. This point also proves the importance of assessing the reading comprehension process appropriately and effectively.

1.1. Problem Statement

The types of assessment techniques to be used is as much important as assessing students' reading comprehension in a reliable and valid way. According to Castillo (2006, cited in Nel, 2011), teachers can make informal assessments based on performance, or they can make formal assessments based on scores, rules, and criteria. This can provide an opportunity to compare students. McAndrews (2008) draws attention to three assessment techniques, especially in assessing primary school students' reading comprehension. He lists these in the form of cloze technique (filling in the blanks), answering open-closed ended or multiple-choice questions, and retelling the read text. These are formally important assessment techniques. In addition, Barrett's taxonomy can be used to assess reading comprehension skills (Göçer, 2018). However, assessing reading comprehension by asking open-closed or multiple-choice questions based on the text is considered the most ideal (Baştuğ et al., 2019) and one of the common assessment techniques (Shi et al., 2018). However, it is also difficult to score answers to open-ended questions that require in-depth understanding and inference. To achieve this, a rubric should be prepared based on the expected skills from the students. The rubric both increases the reliability and validity of the scorings and facilitates the assessment process. In this sense, rubrics are important assessment and evaluation instruments in the assessment of reading comprehension skills. Through rubrics that provide informative feedback on students' strengths and weaknesses (Göçer, 2018), more detailed information can be obtained about students' performance in understanding what they read (Baştuğ et al., 2019). Rubrics have an important place in the assessment of reading comprehension performance. However, there are also some concerns with the rubrics as a performance assessment instrument. As a matter of fact, Romagnono (2001) states that the most important concern about performance assessment is related to the objectivity in the process of scoring the individual's performance and determining the situation. He points out that scoring performance assessment in an objective way is challenging like traditional assessment (fixed response assessment). Therefore, there are some concepts that negatively affect objectivity by making this assessment difficult. Sources that affect or reduce objectivity in the process of performance assessment are explained as rater errors/bias or effects (Farrokhi et al., 2011; Haladyna, 1997; İlhan, 2015; Şata & Karakaya, 2022).

1.2. Related Research

Literature review shows that there are limited number of studies that examine the teachers' bias in assessing the reading comprehension of disadvantaged primary school students. Cosgrove et al. (2000) also draw attention to this deficiency within the scope of studies on the reading comprehension skills of disadvantaged and at-risk children. Likewise, it should be noted that studies in this sense are limited, especially considering that teachers are considered as the sources of rater bias (Mason et al., 2014). On the other hand, Milanowski (2017) questioned in his research whether the low academic performance/scores of disadvantaged students are a bias of the teachers or reflects the real success of the students. Due to the scarcity of studies in the literature on examining rater behaviors in scoring disadvantaged students, this research becomes even more meaningful. It is thought that the present study will contribute to the literature in this respect. This study aims to examine the biases of classroom teachers during the assessment process reading comprehension of primary school fourth grade students who have various disadvantages. The participants of the study were selected from the fourth grade because the participating students were attending the Remedial Education Program and their special situation were already confirmed by the Counseling and Research Center. The most important reason for seeking participation in Remedial Education Program is that many disadvantaged students receive education within the scope of this program (Keskin & Üstün, 2020; Ministry of National Education, 2019).

1.3. Research Objectives

Based on these objectives and reasons, answers to the following questions were sought:

1. Do raters show rater bias in assessing the reading comprehension skills of disadvantaged and non-disadvantaged students?

2. Do the criteria of the reading comprehension rubric differ for disadvantaged and non-disadvantaged students?

2. Theoretical Framework

Assessment of reading comprehension is important in terms of revealing the quality, success, and performance of the reading comprehension process and giving feedback (Klingner et al., 2007). However, as well as the transparency and fairness of the assessment process, especially done by teachers, the measurements obtained from the relevant tools must also be reliable and valid. Likewise, it is known that the target population and its characteristics are also a factor affecting the assessment process (Bergh, 2010). Dovidio et al. (2002) also state that this situation may negatively affect the success level of students. Therefore, teachers need to know the importance of assessing reading comprehension in the classroom (Afflerbach & Cho, 2011; Pressly, 2001), consider the mentioned dynamics and principles, and have effective communication skills. In fact, the role of the teacher and assessment tools in this context is very important (Kenny & Chekaluk, 1993). On the other hand, Antoniazzi et al. (2010) point out that the development of these tools by teachers is especially important in identifying the reading problems of disadvantaged and at-risk children because one of the most important purposes of reading and assessing reading comprehension is to identify the problems experienced by disadvantaged or at-risk children (Carlisle & Rice, 2004). Therefore, the responsibility of identifying students who need support in the classroom and having information about them makes the assessment process more sensitive and critical (Nel, 2011).

The teacher's role in the classroom is crucial, considering that disadvantaged and at-risk children who need support are in many ways backward and especially deprived of family support (Dewulf et al., 2020). In addition, it should be noted that the following groups of children are considered to be disadvantaged students and their language skills are usually at an insufficient level: children with low socio-economic status, ethnic minority (van Ewijk & Slegers, 2010), working in agriculture, migrant-refugee-asylum seeker, exposed to abuse, diagnosed with special learning disabilities by Counseling and Research Centers (CRC), substance abusers, and enrollees of the Primary School Remedial Education Program (Ministry of National Education, 2019; Taneri, 2019). This has a negative effect on reading (OECD, 2010) and most importantly, on reading comprehension (Eivers et al., 2004; Phillips & Lonigan 2005; Schacter & Jo, 2005; Van Keer, 2004). Thus, disadvantaged and at-risk children have an extra difficulty in reading and reading comprehension skills (Cosgrove et al., 2000; Kyriakides, 2000; Lyons et al., 2013) and special care has to be taken to assess their reading performance. It becomes more important for teachers to assess reading performance effectively and accurately, as poor reading and comprehension skills in disadvantaged and at-risk students make them poor readers and affect their performance in other subjects and even their overall academic success (Arnold & Doctoroff, 2003; Bolat, 2021; Crawford et al., 2001).

3. Method

3.1. Research Design

The study was designed as a descriptive study as it is aimed at examining the rater errors/bias, which has a significant effect on the reliability and validity of the measurements in the process of assessing the reading comprehension skills of disadvantaged primary school students. Many-facet Rasch measurement model was used in this study considering its compatibility with the nature of the research data, and a fully crossed pattern was used because of crossing all the surfaces in the research with each other. There are four surfaces in the study: students, raters, student's status (disadvantaged/non-disadvantaged) and assessment criteria. Both surface measurements and interactions of surfaces were considered (Sayın & Şata, 2022).

3.2. Participant

The participants consist of a total of 48 fourth grade students, 13 of whom are disadvantaged and 35 of whom are non-disadvantaged studying in primary schools in a city center in the

Eastern Anatolia region. The research sample was selected by criterion sampling. The disadvantaged participants have the following special situations: weak readers participating Remedial Education Program, children who are seasonal agricultural workers, those who have reading and writing difficulties, foreign nationals, those who received a mild mental disability report from the CRC, students who receive training in the support training room, etc. In particular, the existence of these disadvantage groups within the scope of Remedial Education Program is also found in the research conducted by Keskin and Üstün (2020). In addition, three teachers who were active classroom teachers were raters in the study. All the teachers have nine years of experience, and one is male and two are female.

3.3. Data Collection Tools

'Reading Comprehension Test' and analytical rubric were used to assess reading comprehension in this study as data collection tools. For the reading comprehension test, the narrative text named "Game and Friendship" in the textbook prepared by the Ministry of Education was used. The reading comprehension rubric developed by Baştuğ et al., (2019) was used to measure the reading comprehension skills of the students. The relevant measurement tool consists of seven criteria, graded in four.

Exploratory factor analysis and McDonald ω coefficient were employed to prove the measurements' reliability gathered from the reading comprehension rubric and to validate the inferences drawn from the measurements. The fulfillment of the exploratory factor analysis assumptions was examined, as a result it was investigated that the assumptions were accurate (KMO value was .867 for the relevant data set, while the Bartlett sphericity test was significant; all the criteria of the rubric were normally distributed; there was no missing or extreme value). While performing EFA, the average of the scores given by three raters to the reading comprehension skills of 48 students was taken. There are different views on determining the appropriate sample size for EFA. However, in their simulation study, Guadagnoli & Velicer (1988) emphasized that variables' factor loadings are critical rather than sample size in EFA. According to this study, variables with a high factor load would produce consistent results regardless of the number of variables and sample size of less than 50 people (Guadagnoli & Velicer, 1988). The data analysis revealed that the rubric was collected under a single factor and explained 60.67% of the change in student achievement (The factor load of each criterion in the measurement tool is as follows, respectively; .673; .826; .905; .755; .734; .749 and .788.). The scatter diagram obtained as a result of EFA is given in figure 1.

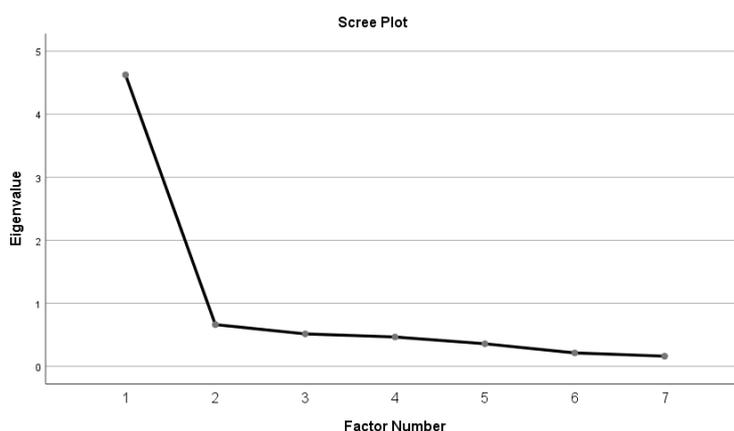


Figure 1. Scatter diagram for the rubric.

In figure 1, based on the ratio between the dominant factor and the second highest factor, it can be seen that it is approximately five times. In this context, because all criteria show a high factor load under one single factor and that the dominant factor has a higher value than other factors has been evaluated as a proof that the measurement tool has a one-dimensional structure.

The reliability coefficient (ω) suggested by McDonald (1999) was administered to bring evidence regarding the measurements' reliability that was obtained after the evidence for the construct validity of the measurement tool was collected. Since the factor loads of the variables are different from each other in this study, the McDonald ω coefficient was preferred as it is aimed to obtain more consistent estimations in such measurements (Osburn, 2000). As a result of the analysis, McDonald ω coefficient was found to be .913 (95% Confidence Interval: .878 - .935).

3.4. Data Collection Process

The data were collected at the end of the 2020-2021 academic year, in May. Before collecting the data, schools where disadvantaged groups continue their education were investigated. Opinions of the experts were taken for this step. The following experts contributed to the mentioned step: provincial deputy director of national education, department chief in the provincial directorate of national education, teachers working in the CRC, and school principals. Afterwards, short interviews were administered with the teachers of the classrooms where the data would be collected by going to the schools. The teachers were informed that the students' reading comprehension levels would be assessed. Attention was drawn to the importance of teachers' contributions to the research. The text named "Game and Friendship" was shared with the teachers and they were asked to apply them in the classroom. Information on rater bias was not shared with the teachers. This measure was taken to prevent the teachers from being biased towards disadvantaged groups in assessing students' reading comprehension scores. It was also specifically stated to the teachers that these assessments are not/would not be part of a rating or a comparison with other classes/schools. It was observed that the students read the text and answered the questions in one lesson hour (40 minutes). The teachers were informed about scoring the comprehension questions regarding the text that was applied simultaneously to the A, B and C branches of the fourth grade and how to use the rubric. Teachers were given two days to assess their students in their classes. After two days, the researcher shared the forms that the teachers submitted among all the teachers. These forms contained the applied text and the comprehension questions. In other words, teachers were asked to assess students in other branches as well and they were given two extra days for this duty. The teachers handed the forms to the researcher at the end. The researcher asked each teacher to assess the reading comprehension of the students of two classes outside of their own class for a third time to complete the process of assessment. Finally, the teachers assessed all the students in A, B and C branches. The opinions of the four experts were taken regarding the rubric before implementing it. Two of the experts were measurement and evaluation specialists and two of them were from the field of Turkish language teaching who specialized in literacy in Turkish language. The experts expressed positive opinions about the use, usefulness, validity, and the items of the rubric.

3.5. Data Analysis

Many-facet Rasch measurement (MFRM) was used to analyze the data. The FACET package program (version 3.70.1) was applied for analyzes. MFRM allows the placement of numerous sources of variability (including rater, item, task, individual, and time) on a single evenly spaced scale (Linacre, 1993). MFRM is another name for facets models (Eckes, 2015). An important point in the MFRM analysis is that it allows the interaction of the sources of variability to be examined in detail (Kassim, 2007). Independent variables (e.g., peer evaluation, status determination criteria, open-ended items) are seen as dependent variables as a result of the logistic transformation of the log odds ratios.

The study's model was established as follows because a group of raters evaluated the reading comprehension abilities of a group of students:

$$\log\left(\frac{P_{bkpx}}{P_{bkpx-1}}\right) = \theta_b - \beta_k - \alpha_p - \tau_x \quad (1)$$

P_{bkpx} = the probability that rater p gives a score of x to the criterion k of student b

P_{bkpx-1} = the probability that rater p gives a score of $x-1$ to the criterion k of student b

θ_b = proficiency level of student b,

β_k = difficulty level of the criterion k,

α_p = the degree of severity of the rater p,

τ_x = difficulty of getting a score of x instead of a score of x-1,

Also, since it was also aimed to examine the interactions of facets with each other, another research model was used:

$$\log\left(\frac{P_{bkpx}}{P_{bkpx-1}}\right) = \theta_b - \beta_k - \alpha_p - \tau_x - I_{pb} \quad (2)$$

The term I in Equation 2 refers to the interaction term between facets. In MFRM, identifying rater mistakes is significantly influenced by the interaction (bias) index (Engelhard, 2002; Linacre, 2017).

For the measurements obtained from MFRM to be unbiased and consistent, some assumptions must be satisfied. These assumptions are unidimensionality, local independence, and model data fit. Since it was shown in the data collection tools section that the measurement tool had a one-dimensional structure, the first assumption was met. Since providing unidimensionality indicates that local independence is also achieved, this assumption was met. Finally, standardized residuals were examined for model data fit. Standardized residuals outside the 2 range should not account for more than 5% of the total number of observations for model data fit, and standardized residuals outside the 3 range should not account for more than 1% of the entire data number (Linacre, 2017). The model-data fit was satisfactory (total number of observations $3 \times 7 \times 48 = 1\,008$) because 44 (4.37%) values were in the range of 2 and 3 (0.30%) values were in the range of 3.

4. Findings

First, whether the reading comprehension levels of the disadvantaged and non-disadvantaged students differed statistically was investigated. Findings were presented in Table 1.

Table 1. Measurement report on the difference between students' reading comprehension levels according to student status

Student types	Observed Mean	Logit measure	Standard error of measure	Infit	Outfit
Disadvantaged	3.08	-0.21	0.09	0.97	1.01
Non-disadvantaged	2.77	+0.21	0.06	1.02	1.00
Mean	2.92	0.00	.08	.99	1.00
Standard deviation	0.22	0.29	.03	0.03	0.00

Model, Sample: RMSE = .08 Standard deviation = 0.28

Separation ratio =3.58 Separation index = 5.10 Reliability of separation index = .93

Model, Fixed (all same) chi-square = 13.8 df = 1 p = .00

As seen in Table 1, the proficiency levels of reading comprehension skills of disadvantaged and non-disadvantaged students had a statistically significant difference (χ^2 (df) = 13.8 (1); $p < .05$). The separation ratio, the separation index and the reliability of the separation index calculated for the difference in the reading comprehension levels of both student groups were high. Regarding the logit values, while the disadvantaged students had a -0.21 value, the non-disadvantaged students had +0.21. Besides, the observed mean showed that the non-disadvantaged group had a mean of 3.08 out of four, while the disadvantaged group had a mean of 2.77.

After determining that the reading comprehension levels of the students were different from each other, the teachers' display of scoring errors in the process of assessing the students' reading comprehension levels were examined. For this purpose, the interactions of *student types x rater* were examined. Table 2 presented these findings.

Table 2. Statistically Significant Interactions Between the Raters and Student Types

Rater	Student types	Observed Score	Expected Score	Bias (logit)	Standard Error	<i>p</i>
R1	Disadvantaged	259	243.51	0.41	.16	.014*
R2	Non-disadvantaged	804	786.82	0.19	.11	.077
R3	Disadvantaged	249	247.32	0.04	.16	.784
R3	Non-disadvantaged	741	742.68	-0.02	.10	.867
R1	Non-disadvantaged	717	732.49	-0.15	.10	.125
R2	Disadvantaged	247	264.18	-0.46	.16	.006*

Fixed chi-square = 20.0 df = 6 *p* = .000

Not. **p* < .05

As seen in Table 2, teachers who were raters during the assessment process of reading comprehension skills of disadvantaged and non-disadvantaged students made biased scores (χ^2 (df) = 20.0 (6); *p* < .05). The first rater gave more generous ratings to the disadvantaged students, and this difference was statistically significant (*p* < .05). However, the second rater gave lower scores to the disadvantaged students and this difference was statistically significant (*p* < .05). The third rater assessed both groups of students in an unbiased way.

A possible source of disadvantaged students' lower reading comprehension skills was the possibility that the assessment criteria in the measurement tool would be more difficult for these students. In this context, *student type x criterion* interactions were also examined. Table 3 presented the statistics related to these interactions.

Table 3. Statistically Significant Interactions Between the Student Types and Criteria

Rater	Student types	Observed Score	Expected Score	Bias (logit)	Standard Error	<i>p</i>
C5	Disadvantaged	123	116.47	0.44	0.26	.104
C6	Disadvantaged	106	103.19	0.17	0.25	.490
C7	Non-disadvantaged	327	322.46	0.11	0.15	.488
C4	Non-disadvantaged	336	333.31	0.07	0.16	.676
C2	Non-disadvantaged	323	321.01	0.05	0.15	.761
C3	Disadvantaged	108	107.81	0.01	0.25	.963
C1	Non-disadvantaged	306	305.70	0.01	0.15	.963
C3	Non-disadvantaged	323	323.18	0.00	0.15	.978
C1	Disadvantaged	101	101.30	-0.02	0.25	.941
C6	Non-disadvantaged	308	310.81	-0.06	0.15	.672
C2	Disadvantaged	105	106.99	-0.12	0.25	.622
C5	Non-disadvantaged	339	345.52	-0.17	0.16	.295
C4	Disadvantaged	109	111.68	-0.17	0.25	.504

C7	Disadvantaged	103	107.54	-0.28	0.25	.265
----	---------------	-----	--------	-------	------	------

Fixed chi-square = 7.3 df = 14 p = .920

Not. *p < .05

Regarding, Table 3, the assessment criteria in the measurement tool were similar for both student groups, and there was no statistically significant difference (χ^2 (df) = 7.3 (14); $p > .05$). When it was examined in terms of each criterion, it was observed that all criteria worked the same for disadvantaged and non-disadvantaged students ($p > .05$). In other words, there was no differentiated item function in the criteria in the rubric.

5. Discussion

The present paper aimed to examine classroom teachers' bias in assessing the reading comprehension skills of disadvantaged and non-disadvantaged primary school students. In other words, 'whether the teachers were under the influence of the students' disadvantageous situations when the disadvantaged students got low or high scores as a result of their reading comprehension performance' should be examined. These inquiries also provided the opportunity to compare the reading comprehension performances of disadvantaged students and non-disadvantaged students. Finally, it was revealed whether the rubric and its criteria, which were used as an assessment scale, were understood and worked in the same way for both groups of students. While assessing the reading comprehension skills of both groups of students, it is important to discuss the results of the inquiries provided by the questions to which answers are sought in the context of teachers' bias.

The study reported that the classroom teachers were biased in the scores they gave in assessing the reading comprehension skills of the disadvantaged students. While one of the classroom teachers gave more points while assessing the reading comprehension skills of disadvantaged students, the other teacher gave fewer points. The third-grade teacher, on the other hand, acted impartially in the evaluation. It is necessary to discuss in detail the level of bias in classroom teachers' scores while assessing the reading comprehension skills of disadvantaged students and its reasons. It is very difficult to make a good and effective assessment for disadvantaged students (Milanowski, 2017). Such differences in scores may also reflect teaching practices for disadvantaged students (AERA, 2014). Figure 2 shows the cycle of teachers' scoring biases for disadvantaged students.

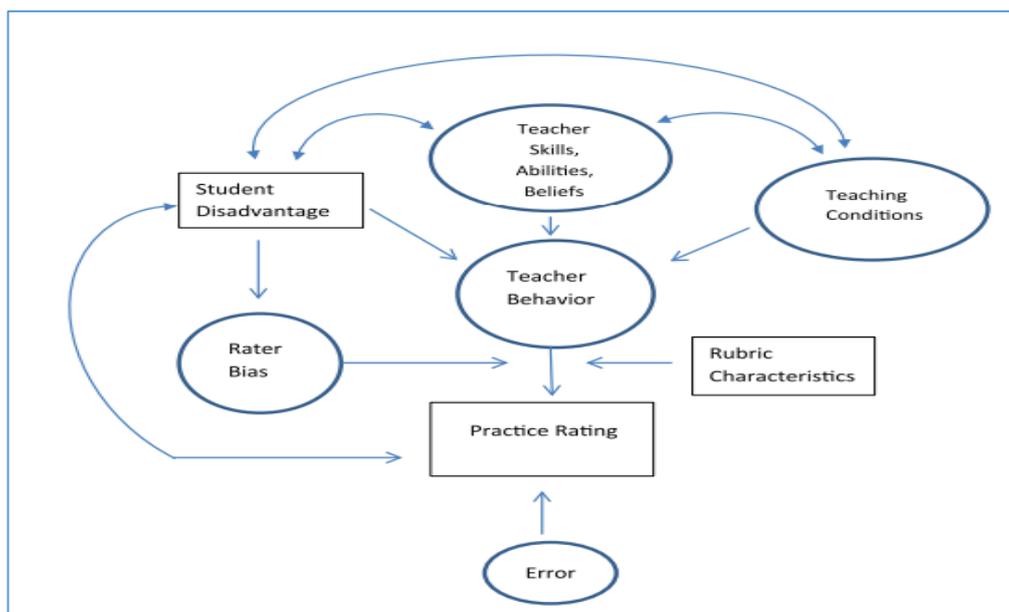


Figure 2. Factors affecting the biases in teachers' scoring of students with severe or low disadvantage (Source: Milanowski, 2017).

Regarding Figure 2, the disadvantage of students was determinant in teachers' skills, beliefs, and abilities, and this teaching process affected teacher behavior. Similarly, being disadvantaged was associated with rater bias. This situation can be interpreted as teachers may show bias in the scores of assessing the reading comprehension skills of disadvantaged students. Supporting this finding, Milanowski (2017) states that this situation can be explained by the teacher's lack and inadequacy in teaching skills (Clotfelter et al., 2005) and the limited number of teaching materials that facilitate teaching. This finding is consistent with the results of various studies. Some studies acknowledge that teachers are biased in scoring different applications in areas such as reading, writing, and mathematics to students in various disadvantaged groups who cannot benefit from effective teaching (Isenberg, 2013; Quinn, 2020; Sass et al., 2012). In this context, McDonald (1998) emphasizes that teachers are biased in scoring because of their low expectations for disadvantaged students. As a result of this bias, students are given lower scores. Whitehead (2007) states that the problems that arise due to teachers' negative views and prejudices towards disadvantaged students (Kellaghan & Fontes, 1989) are not limited to the decline in academic achievements such as reading. It is also stated that biased behavior also affects the school attendance of students, especially those who are socio-economically disadvantaged, in the long run. The biases of teachers in assessment scores of disadvantaged students and students' performances are not limited to reading and reading comprehension. This attitude of teachers also affects grades in other courses. For example, Strambler and Weinstein (2010) emphasize that the negative attitude of teachers towards disadvantaged students affects other lessons and is reflected in the grades in the mathematics lesson. Aytaş and Kardaş (2014) draw attention to a different situation. Accordingly, the assessment part is incomplete because disadvantaged students cannot fully answer the questions. It can be said that this situation may create a bias effect, albeit partially, in scoring the reading comprehension skills of disadvantaged students. It is possible to present a different perspective on the fact that teachers have a significant impact on the reading success and evaluation of students in different disadvantaged groups. Such a result is reported by Peterson et al., (2016). According to their findings, the teacher does not have a prejudiced and biased attitude towards disadvantaged groups. Therefore, they underline that since there is no such approach, it is not possible to mention that this situation affects reading success. This can be considered as an important reference in terms of revealing the effect of the presence or absence of teachers' prejudiced and biased attitudes towards disadvantaged students on reading performance. Moreover, Mortimore and Blackstone (1982) discuss why teachers act biased and prejudiced against disadvantaged students from a different perspective. They state that teachers do not have any expectations for students. This point of view is decisive in affecting the bias in students' scores. Stating that traditional assessment approaches are a problem for disadvantaged students, Natriello et al. (1991) advocate that more sensitive assessment methods should be developed for these students. They attribute the biased behaviors to the inadequacy and lack of alternatives in this regard. Pointing out that the teacher may face situations of prejudice and bias while assessing the performance of disadvantaged students, they express the necessity and importance of these alternatives.

Findings showed that the reading comprehension skills of the non-disadvantaged students were higher than the disadvantaged students. Although this was due to the bias of the two teachers in scoring, the fact that one rater acted impartially was a matter of discussion. Various studies reported similar findings, indicating that the reading comprehension skills of the disadvantaged students were at a lower level in terms of their reading performance when compared to the non-disadvantaged students (Barton & Coley, 2009; McDonald, 1998). Lyons et al., (2013) state that disadvantaged primary school students are more unsuccessful in reading than their non-disadvantaged peers. They also emphasize that disadvantaged students are at risk in terms of reading performance. Besides, they state that this problem is more evident especially in students who are in the socioeconomically disadvantaged group. According to Eivers et al. (2005), there is a direct relationship between being disadvantaged and reading difficulties. In this context, they draw attention to the fact that students in disadvantaged groups have more difficulties in reading and have lower reading performance scores when compared to non-disadvantaged groups due to their various characteristics. They acknowledge that the reading performance of boys is lower than that of girls. Kellaghan et al.

(1995) emphasize that disadvantaged students are at a lower level in reading success than other students. The low reading achievement of disadvantaged students is a serious problem. It can be said that reading is an important factor and a predictor of general success. Therefore, it is worth considering this problem which disadvantaged groups have. Similar thoughts are also expressed by Juel (1991). According to Kennedy (2008), some interventions are needed to minimize the problems of disadvantaged students regarding this issue. It is important to provide teachers with opportunities to apply new and effective teaching approaches and to provide professional development support. Likewise, it is necessary to ensure the use of activities that will encourage children's creativity and participation. Most importantly, there is a need to carry out comprehensive and well-equipped studies for the schools included in this scope. The OECD (2017) also reports similar recommendations for improving the reading performance of disadvantaged students. Noting that good socio-economic conditions are effective in providing reading opportunities for parents to their children, the report emphasizes this deficiency in disadvantaged students. It is understood that unequal socioeconomic conditions cause differences in reading performance between the two groups. Supplementary education is needed to minimize these differences between disadvantaged students and non-disadvantaged students. Yue et al., (2018) advocate that the differences between the groups decrease with the support and reinforcement education to be given to non-disadvantaged students.

The present study showed that the assessment criteria in the measurement tool were similar for both student groups, indicating that there was no statistically significant difference. This can be interpreted as the rubric items were not effective in keeping the reading comprehension skills of disadvantaged students at a lower level. In other words, disadvantaged students and non-disadvantaged students understood the rubrics in the same way. Mongkuo and Mongkuo (2017) also tried to demonstrate the validity of the test by applying it to disadvantaged students and regular students. They found no significant difference between the groups, drawing attention to the validity level of the test. In other words, the disadvantaged students did not have a problem in understanding and scoring the performance test items. It can be said that there are no different results in terms of disadvantaged and non-disadvantaged students under equal conditions. This is also true for classroom environments with alternative assessment tools, methods, and practices. This was tested in a classroom with an active teacher. In their study, Flanagan and Addy (2019) found that the same level of questions was understood in the final questions applied between disadvantaged and non-disadvantaged students, so there was no difference between the groups.

6. Conclusion

The study reported that the classroom teachers were biased in the scores they gave in assessing the reading comprehension skills of the disadvantaged students. Findings showed that the reading comprehension skills of the non-disadvantaged students were higher than the disadvantaged students. Although this was due to the bias of the two teachers in scoring, the fact that one rater acted impartially was a matter of discussion. The present study showed that the assessment criteria in the measurement tool were similar for both student groups, indicating that there was no statistically significant difference. This can be interpreted as the rubric items were not effective in keeping the reading comprehension skills of disadvantaged students at a lower level. In other words, disadvantaged students and non-disadvantaged students understood the rubrics in the same way.

Limitation

Research data is limited to data collected from three raters. Data were also collected from a school where disadvantaged primary school students are heavily educated. The study data carried out with the quantitative research method are also limited to quantitative data tools. Therefore, it should be noted that the measurement tools are limited to the reading comprehension test and the rating scale.

Recommendation

Based on these findings, the following suggestions can be made: It is necessary to develop reliable and valid measurement tools that could assess the reading and other skills of disadvantaged and at-risk students. In-service training should be given to classroom teachers to prevent teachers from being biased in their positive/negative scoring or to minimize the effect of this problem. Besides, teachers' assessment tools and their scoring should also be supervised. Inclusive assessment techniques should be made a part of the classroom by fulfilling the requirements of inclusive education. Due to linguistic differences, supportive training should be given to disadvantaged students. Compensation training should be given to children who are seasonal agricultural workers and those with special learning difficulties. Efforts should be made to bring disadvantaged students to the same level as other students through extra education. Further studies may investigate whether classroom teachers show similar bias when assessing the exams and written expression skills of disadvantaged primary school students. Moreover, the negative consequences of classroom teachers' biases in assessing disadvantaged students, especially reading skills, as raters can be examined

Acknowledgments

This study is not supported by a foundation.

Conflict of Interest

The researchers declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Afflerbach, P., & Cho, B. (2011). The classroom assessment of reading. M. Kamil, P. Pearson, E. Moje, & P. Afflerbach, (Eds.), in *Handbook of reading research*. New York. <https://doi.org/10.4324/9780203840412>
- Akyol, H. (2019). *Turkish primary reading and writing teaching*. Pegem Publishing. <https://doi.org/10.14527/9786050370171>
- American Educational Researchparticularation, American Psychological Association, & National Council on Measurement in Education (AERA). (2014). *Standards for educational and psychological testing*. Washington. <https://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition>
- Antoniazzi, D., Snow, P., & Swift, V. D. (2010). Teacher identification of children at risk for language impairment in the first year of school. *International Journal of Speech-Language Pathology*, 12(3), 244-252. <https://doi.org/10.3109/17549500903104447>
- Arnold, D. H., & Doctoroff, G. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, 54, 517-545. <https://doi.org/10.1146/annurev.psych.54.111301.145442>
- Aytaş, G., & Kardaş, D. (2014). Dezavantajlı öğrencilerin okuma anlama gelişimleri. *Turkish Studies*, 9(1), 79-92. <https://doi.org/10.7827/TurkishStudies.7493>
- Baddour, J., Hipfner-Boucher, K., Janaideh, R. A., Paradis, J., Gottardo, A., & Chen, X. (2019). A Snapshot of progress: The development of English and Arabic language and literacy skills of Syrian refugee children and youth in Canada. *Selected presentations form the 22nd Metropolis Canada Conference* (pp. 52-72).
- Barton, P., & Coley, R. (2009, 10 01). *Parsing the Achievement Gap II. Educational Testing Service*. <http://www.ets.org/Media/Research/pdf/PICPARSINGII.pdf>

- Baştuğ, M., Hiğde, A., Çam, E., Örs, E., & Efe, P. (2019). *Developing reading comprehension skills: Strategies, techniques, applications*. Pegem Publishing. <https://doi.org/10.14527/9786052415504>
- Bergh, V. L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47(2), 497-527. <https://doi.org/10.3102/0002831209353594>
- Bolat, Y. (2021). Primary school teachers' views on Syrian Students' Turkish and math skills and the confronted challenges. *International Journal of Modern Education Studies*, 5(1), 92-117. <http://dx.doi.org/10.51383/ijonmes.2021.84>
- Carlisle, J., & Rice, M. (2004). Assessment of reading comprehension. A. Stone, E. Silliman, B. Ehren, & K. Apel, (Eds.), In *Handbook of language and literacy* (pp. 521-555). Guilford Publications. <https://www.guilford.com/books/Handbook-of-Language-and-Literacy/Stone-Silliman-Ehren-Wallach/9781462527489/reviews>
- Clotfelter, C., Ladd, H. F., & Vigdor, J. L. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, 24(4), 377-392. <https://doi.org/10.1016/j.econedurev.2004.06.008>
- Cosgrove, J., Kellaghan, T., Forde, P., & Morgan, M. (2000). *The 1998 national assessment of English reading*. Educational Research Centre. <https://www.erc.ie/2000/03/01/publications-2000/>
- Crawford, L., Tindal, G., & Steiber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*, 7(4), 303-323. https://doi.org/10.1207/S15326977EA0704_04
- Dewulf, L., Braak, J. V., & Houtte, M. (2020). Examining reading comprehension in disadvantaged segregated classes. The role of class composition, teacher trust, and teaching-learning strategies. *Research Papers in Education*, 1-22. <https://doi.org/10.1080/02671522.2020.1864766>
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62-68. <https://doi.org/10.1037/0022-3514.82.1.62>
- Eckes, T. (2015). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments*. Peterlang. <https://doi.org/10.3726/978-3-653-04844-5>
- Eivers, E., Shiel, G., & Shortt, F. (2004). *Reading literacy in disadvantaged primary schools*. Educational Research Centre. <https://library.ncirl.ie/items/15358>
- Engelhard, G. (2002). Monitoring raters in performance assessments. In G. Tindal and T. Haladyna (Eds.), *Large-scale assessment programs for ALL students: Development, implementation, and analysis* (pp. 261-287). Routledge. <https://doi.org/10.4324/9781410605115>
- Esfandiari, R. (2015). Rater errors among peer-assessors: applying the many-facet Rasch measurement model. *Iranian Journal of Applied Linguistics*, 18(2), 77-107. <https://doi.org/10.18869/acadpub.ijal.18.2.77>
- Farrokhi, F., Esfandiari, R., & Vaez Dalili, M. (2011). Applying the many-facet Rasch model to detect centrality in self-assessment, peer-assessment and teacher assessment. *World Applied Sciences Journal*, 15(11), 70-77. [http://www.idosi.org/wasj/wasj15\(IPLL\)11/12.pdf](http://www.idosi.org/wasj/wasj15(IPLL)11/12.pdf)
- Flanagan, K. M., & Addy, H. (2019, 10 02). *Introverts are not disadvantaged in group-based active learning classrooms*. <https://files.eric.ed.gov/fulltext/EJ1223951.pdf>
- Göçer, A. (2018). *Process and result evaluation methods and tools: Measurement and evaluation in Turkish education*. Pegem Publishing.

<https://doi.org/10.14527/9786053648079>

- Grabe, W., & Stoller, F. L. (2002). *Teaching and researching reading*. Pearson Education Longman. <https://doi.org/10.4324/9781315833743>
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265-275. <https://doi.org/10.1037/0033-2909.103.2.265>
- Habib, M. (2016). Assessment of reading comprehension. *Revista Romaneasca pentru Educatie Multidimensionala*, 8(1), 125-147. <http://dx.doi.org/10.18662/rrem/2016.0801.08>
- Haladyna, T. M. (1997). *Writing test items in order to evaluate higher-order thinking*. Allyn & Bacon.
- İlhan, M. (2015). *The identification of rater effects on open-ended math questions rated through standard rubrics and rubrics based on the SOLO taxonomy in reference to the many facet rasch model. (Unpublished doctoral dissertation)*. <https://tez.yok.gov.tr>
- Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). *Access to effective teaching for disadvantaged students*. National Center for Education Evaluation and Regional Assistance. <https://files.eric.ed.gov/fulltext/ED544345.pdf>
- Juel, C. (1991). Beginning reading. In R. Barr, M. Kamil, P. Mosenthal, & P. Pearson (Eds.), *Handbook of Reading Research* (pp. 759-788). Routledge. <https://www.routledge.com/Handbook-of-Reading-Research/Pearson-Barr-Kamil-Mosenthal-Barr/p/book/9780805841503>
- Kassim, N. L. A (2007). *Exploring rater judging behaviour using the many-facet Rasch model*. Paper Presented in the Second Biennial International Conference on Teaching and Learning of English in Asia: Exploring New Frontiers (TELiA2), Universiti Utara, Malaysia. <http://repo.uum.edu.my/3212/>
- Kellaghan, T., & Fontes, P. (1989). Relative bias in teacher judgements and standardized tests in the identification of literacy problems. *Irish Journal of Education*, 23(2), 40-50. <https://www.erc.ie/documents/vol23chp3.pdf>
- Kellaghan, T., Weir, S., Ó hUallacháin, S., & Morgan, M. (1995). *Educational disadvantage in Ireland*. (No. 20). Combat Poverty Agency. [https://books.google.com.tr/books?hl=tr&lr=&id=7raCQTJrWVsC&oi=fnd&pg=PA3&dq=Kellaghan,+T.,+Weir,+S.,+%C3%93+hUallach%C3%A1in,+S.,+%26+Morgan,+M.,+\(1995\).+Educational+disadvantage+in+Ireland.+Department+of+Education.&ots=Z4SG0naVPw&sig=QXqxSbwznlqHXwogGX3Ar8FpU0o&redir_esc=y#v=onepage&q&f=false](https://books.google.com.tr/books?hl=tr&lr=&id=7raCQTJrWVsC&oi=fnd&pg=PA3&dq=Kellaghan,+T.,+Weir,+S.,+%C3%93+hUallach%C3%A1in,+S.,+%26+Morgan,+M.,+(1995).+Educational+disadvantage+in+Ireland.+Department+of+Education.&ots=Z4SG0naVPw&sig=QXqxSbwznlqHXwogGX3Ar8FpU0o&redir_esc=y#v=onepage&q&f=false)
- Kennedy, E. (2008). *Improving literacy achievement in a disadvantaged primary school: Empowering classroom teachers through professional development (Unpublished doctoral dissertation)*. Dublin, Ireland.
- Kenny, D., & Chekaluck, E. (1993). Early reading performance: A comparison of teacher-based assessments. *Journal of Learning Disabilities*, 26, 227-236. <https://doi.org/10.1177/002221949302600403>
- Klingner, J. K., Vaughn, S., & Boardman, A. (2007). *Teaching reading comprehension to students with learning difficulties*. Guilford Publications. <https://www.guilford.com/books/Teaching-Reading-Comprehension-to-Students-with-Learning-Difficulties/Klingner-Vaughn-Boardman/9781462517374>
- Kyriakides, L. R., Campbell, J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of creemers comprehensive model of educational effectiveness. *School Effectiveness and School Improvement*, 11(4), 501-529. <https://doi.org/10.1076/sesi.11.4.501.3560>

- Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transaction*, 7(1), 283-284. <https://www.rasch.org/rmt/rmt71h.htm>.
- Linacre, J. M. (2017). *A user's guide to FACETS: Rasch-model computer programs*. [computer software manual]. www.winsteps.com
- Lyons, R., Fives, A., Kearns, N., Canavan, J., Devaney, C., & Eaton, P. (2013). Exploring the utility of the simple view of reading in Irish children attending schools in areas designated as socially disadvantaged. *Journal of Education and Training Studies*, 1(1), 204-220. <http://dx.doi.org/10.11114/jets.v1i1.107>
- Mason, B. A., Gunersel, A. B., & Ney, E. A. (2014). Cultural and ethnic bias in teacher ratings of behavior: A criterion-focused review. *Psychology in the Schools*, 51(10), 1017-1030. <https://doi.org/10.1002/pits.21800>
- McAndrews, S. L. (2008). *Diagnostic literacy assessments and instructional strategies*. International Reading Assoc. <https://searchworks.stanford.edu/view/7753115>
- McDonald, E. (1998). Reading Achievement and Educational Disadvantage. *Irish Educational Studies*, 17(1), 208-221. <http://dx.doi.org/10.1080/0332331980170119>
- McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum Associates <https://doi.org/10.4324/9781410601087>
- Milanowski, A. (2017). Lower performance evaluation practice ratings for teachers of disadvantaged students: Bias or reflection of reality? *AERA Open*, 3(1), 1-16. <https://doi.org/10.1177/2332858416685550>
- Mongkuo, M. Y., & Mongkuo, M. Y. (2017). Testing the factorial equivalence of the collegiate learning assessment performance task diagnostic instrument across lower class and upper class predominantly black college students. *International Journal of Education and Practice*, 5(6), 95-103. <https://doi.org/10.18488/journal.61.2017.56.95.103>
- Mortimore, J., & Blackstone, T. (1982). *Disadvantage and education*. Heinemann.
- Natriello, G., Pallas, A. M., & Riehl, C. (1991). *Creating more responsive student evaluation systems for disadvantaged student*. <https://files.eric.ed.gov/fulltext/ED333042.pdf>
- Nel, C. (2011). Classroom assessment of reading comprehension: How are preservice foundation phase teachers being prepared? *Per Linguam*, 27(2), 41-66. <http://dx.doi.org/10.5785/27-2-107>
- OECD. (2012). *PISA 2012 results: Excellence through equity*. OECD Publishing. <https://doi.org/10.1787/19963777>
- OECD. (2017). *Could learning strategies reduce the performance gap between advantaged and disadvantaged students?* OECD Publishing. <https://doi.org/10.1787/5k4220w36z25-en>
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological methods*, 5(3), 343. <http://dx.doi.org/10.1037/1082-989X.5.3.343>
- Peterson, E. R., Davies, C. R., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, 42, 123-140. <http://dx.doi.org/10.1016/j.learninstruc.2016.01.010>
- Phillips, B., & Lonigan, C. (2005). Social correlates of emergent literacy. In M. Snowling, & C. Hulme, (Eds.), *Handbooks of developmental psychology. The science of reading: A handbook* (pp. 173-187). Malden. <https://www.wiley.com/en-ie/The+Science+of+Reading:+A+Handbook,+2nd+Edition-p-9781119705130>
- Pressly, M. (2001). *Effective beginning reading instruction. Executive summary and paper commissioned by the National Reading Conference*.
- Quinn, D. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role

- of grading scales. *Educational Evaluation and Policy Analysis*, 1-64. <https://doi.org/10.3102/0162373720932188>
- Report of the National Reading Panel (NRP). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report_t.pdf
- Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Mathematics Teacher*, 94(1), 31-37. <http://peteriljedahl.com/wp-content/uploads/Myth-of-Objectivity2.pdf>
- Sass, T., Hannaway, J., Xu, Z., Figlio, D., & Feng, L. (2012). Value-added of teachers in high-poverty schools and lower-poverty schools. *Journal of Urban Economics*, 72, 104-122. <https://doi.org/10.1016/j.jue.2012.04.004>
- Sayın, A. & Şata, M. (2022). Using Rasch analysis to examine raters' expertise Turkish teacher candidates' competency levels in writing different types of test items. *International Journal of Assessment Tools in Education*, 9(4), 998-1012. <https://doi.org/10.21449/ijate.1058300>
- Schacter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, 28(2), 158-169. <https://doi.org/10.1111/j.1467-9817.2005.00260.x>
- Si, G., Lipper, A., Shubeck, K., Fang, Y., Chen, S., Pavlik, P., . . . Graesser, A. (2018). Exploring an intelligent tutoring system as a conversation-based assessment tool for reading comprehension. *Behaviormetrika*, 1-19. <https://doi.org/10.1007/s41237-018-0065-9>
- Solari, E. J., Grimm, R. P., McIntyre, N. S., & Denton, C. A. (2018). Reading comprehension development in at-risk vs. not at-risk first grade readers: The differential roles of listening comprehension, decoding, and fluency. *Learning and Individual Differences*, 65, 195-206. <https://doi.org/10.1016/j.lindif.2018.06.005>
- Strambler, M. J., & Weinstein, R. S. (2010). Psychological disengagement in elementary school among ethnic minority students. *Journal of Applied Developmental Psychology*, 31, 155-165. <http://dx.doi.org/10.1016/j.appdev.2009.11.006>
- Şata, M. & Karakaya, İ. (2022). Investigating the impact of rater training on rater errors in the process of assessing writing skill. *International Journal of Assessment Tools in Education*, 9(2), 492-514. <https://doi.org/10.21449/ijate.877035>
- Taneri, P. (2019). *Kuramdan uygulamaya kapsayıcı eğitim*. Pegem Publishing. <https://doi.org/10.14527/9786052418345>
- van Ewijk, R., & Slegers, P. (2010). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21(3), 237-265. <https://doi.org/10.1080/09243451003612671>
- Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology*, 74(1), 37-70. <https://doi.org/10.1348/000709904322848815>
- Wesolowski, B. C., Wind, S. A., & Engelhard Jr, G. (2015). Rater fairness in music performance assessment: Evaluating model-data fit and differential rater functioning. *Musicae Scientiae*, 19(2), 147-170. <https://doi.org/10.1177/1029864915589014>
- Whitehead, U. O. (2007). African-American students' perceptions of teacher attitudes on academic achievement and discipline sanctions. *ProQuest Information & Learning*, 68. https://aquila.usm.edu/theses_dissertations/954/

Yusuf Kızıldaş, Mehmet Şata & Fuat Elkonca, *Classroom Teachers' Bias in Assessing Disadvantaged...*

Yue, H., Rico, R. S., Vang, M. K., & Giuffrida, T. A. (2018). Supplemental instruction: Helping disadvantaged students reduce performance gap. *Journal of Developmental Education*, 41(2), 18-25. <https://files.eric.ed.gov/fulltext/EJ1200705.pdf>.