PPSE P121 and P10 calculation method and related issues

Umit Celen (iD)[1,*]

[1]Amasya University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation, Amasya, Turkey

Abstract: This study examined the calculation methods of P121 and P10 scores used in teacher appointments. The statistics regarding the Public Personnel Selection Examination (PPSE) subtests used by Measurement, Selection and Placement Center (MSPC) in 2018, 2019 and 2020 were accessed from the website of the institution. The parameters not published on this webpage were calculated by using the candidates' results. The public openly debates the allegations made by the candidates who took the exam in 2019 that their scores had been miscalculated for various reasons and the examinee scores, in fact, had to be higher. The study was conducted (i) to determine whether such disparity actually existed, (ii) and if so, the reason behind it, (iii) how the differences arising from the parameters in the formula being used to calculate the scores would affect exam takers' scores. In particular, the study identified the issues caused by converting the scores obtained by using different subtests in the same manner in calculating P121without considering an equating method. Based on the examined exam scores for the last three-years, it was concluded that 2019 candidates were disadvantaged in most teaching fields. Based on the findings, it is suggested that (i) the use weighted standard scores instead of P121 and P110, to calculate separate scores for each teaching field is better and (ii) the validity period of such exam scores should be limited to one year.

1. INTRODUCTION

Since 2002, the Ministry of National Education (MoNE) has been utilizing the scores obtained in the Public Personnel Selection Exam (PPSE) held annually by the Measurement Selection and Placement Center (MSPC), to appoint new teachers to its affiliated institutions. Before this exam, the Selection Exam for Civil Servants who would be appointed for the first time (SECS) had been used starting in 1999. Initially, the validity period of the exam scores was determined to be 2 years and it was implemented in this manner until 2013. Between 2013 and 2016, the validity period for the scores was 1 year, but with the change in the regulation published in the Official Gazette dated 15 August 2018, the validity period, which was increased to 2 years again starting with the 2017 exam, was reduced to 1 year again with the change on 7 November 2019.

---

*CONTACT: Ümit Çelen ✉ umitcelen@yahoo.com 🖥 Amasya University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation, Amasya, Turkey

Initially, teacher candidates had to take the General Culture (GC), General Ability (GA) and Educational Sciences (ES) tests from the PPSE tests, however, MSPC announced on 27 February 2013 that Teaching Content Knowledge Test (TCKT) would be given in the fields of Turkish, primary school mathematics, science/science and technology, social studies, Turkish language and literature, geography, mathematics (high school), physics, chemistry, biology, religious culture and ethics, and foreign language (German, French, English). In 2014, counseling and classroom teaching, in 2017 pre-school teaching and in 2019 physical education and religious vocational school vocational teaching were added to these fields. In 2019, the number of questions in TCKT was increased from 50 to 75.

When the scores from two different exams are to be used in the same application, the equivalency of these scores becomes important. It is very difficult for tests consisting of different questions to be completely parallel and to produce the same or similar results for each individual who takes the test and therefore, this cannot be expected (Kan, 2010). This difficulty arises from the limitation of the Classical Test Theory which states that "all test and item statistics obtained are affected by the group to which the test is applied". The statistics of test items depend on the sample and are interpreted depending on the group to which the test is applied. (Embreston & Reise, 2000).

In addition, it can be argued that individuals' abilities may vary depending on the items they respond to and that they can perform differently in tests with different difficulty levels (Hambleton & Swaminathan, 1985). In specific, it is possible to equate the scores of tests applied to different people with different questions by using Item Response Theory framework (Crocker & Algina, 1986) but it is publicly known that such an equating process is not used in the exams conducted by MSPC. Instead, "the method of standardizing the scores by relieving them from the effects of mean and standard deviation" which is the most common method used for comparison (Tekin, 1996; Turgut & Baykul, 2010) is preferred.

In each of the PPSE tests, MSPC takes the difference between candidates' correct numbers of questions, corrected for their lucky guessing, the average of the test. Then divides it by the standard deviation of the test to obtain the z-score and multiplies this score by 10 and adds 50 to it to obtain the t-score. Thus, within the limitations of the Classical Test Theory, the performances of the candidates in all tests become relatively comparable and collectable even though they have been taken from different tests. In the calculation of P121, t-scores are calculated by multiplying the sum of General Culture (GC) and General Ability (GA) test scores by coefficient 0.15, the Educational Sciences (ES) Test score by coefficient 0.2, and the Teaching Content Knowledge Test (TCKT) score by coefficient 0.5 and the Weighted Standard Score (WSS) is calculated. When P10 is calculated, WSS is found by multiplying and adding the scores of the GC, GA and ES tests with the coefficients 0.3, 0.3 and 0.4, respectively. Following this conversion, MSPC uses the following formula to get a score out of 100 for each PPSE score type (Measurment, Selection and Placement Center, 2019).

$$PPSE\ Score = 70\ + \frac{30\ [\ 2\ (WSS - X\ ) - S]}{[\ 2\ (B - X)] - S}$$

Abbreviations
PPSE : Public Personnel Selection Examination
WSS : Weighted Standard Score
X   : Average of the WSS distribution
S   : Standard deviation of the WSS distribution
B   : The highest score in the WSS distribution

Examination of PPSE guidelines shows that this formula was first used in 1999 in the Civil Servants Exam (CSE) manual, which was an exam given before the PPSE (Measurment, Selection and Placement Center, 1999). The formula includes the arithmetic mean and standard deviation from all candidates, as well as the B. Çelen's (2013) study based on 2010 PPSE data presents the effect of this transformation, applied to obtain the highest score 100 in each score type, on candidates' scores points out the possible validity issues in PPSE scoring methods.

Candidates who took the PPSE in 2019 claimed that they encountered an unfair situation as they were appointed together with the candidates who took the exam in 2018 because of the increase in the number of items in the teaching field knowledge test in 2019, the change in the duration of the exam, and the higher level of difficulty in the items. They brought it to the attention of the public that although they were at the top of the rankings, they were not appointed but the candidates who ranked lower in the 2018 exam were appointed. This study aimed to determine whether the professed unfairness in 2018 and 2019 scores really occurred and if there was indeed unfairness in the exam scores, to identify the reason behind it. For this purpose, answers to the following research questions were sought.

1. What are the X, S, and B values used in the P-121 account of the PPSE 2018, 2019 and 2020 exams?

2. According to the calculated parameters, does the P-121 corresponding to the same ASP change in 2018, 2019 and 2020?

3. Does the ASP required to have the score of the last appointed person in an appointment period vary in 2018, 2019 and 2020?

4. Are the assignment percentages of teachers appointed with 2018 and 2019 scores different?

5. What are the X, S, and B values used in the P-110 account of the PPSE 2018, 2019 and 2020 exams?

2. METHOD

In this research in the descriptive survey model, the population is the 2018, 2019 and 2020 PPSE scores. 70 result documents were achieved for each year. Thus, the sample consists of the result documents of 210 candidates.

The arithmetic means and standard deviation values of the PPSE subtests in 2018, 2019 and 2020 were used in this study. These values were taken from the MSPC web page. The formulas in the test manual were used for calculating the WSS of the candidates who took the exams. Based on candidates' correct and incorrect number of answers, corrections for lucky guessing were calculated and z-scores were obtained by taking the difference of these values from the mean and dividing them by the standard deviation. Then, these z-scores were multiplied by 10 and converted into t-scores by adding 50. The WSS were calculated by multiplying the t-scores with the coefficients of the PPSE subtests.

Unlike the 2018 manual, page 46 of the guide published by MSPC for 2019 PPSE included 18 TCKT score types (Measurement, Selection and Placement Centre, 2019). Again, page 23 of the same document specified that "PPSE score distribution will be obtained for each PPSE score type out of 100." When this specification and the PPSE P121 score types added to the guide in 2019 were taken together, there was a perception that P121 would be calculated by using the highest scoring WSS for each teaching field that year. Therefore, while trying to estimate the parameters of the exam, different calculations were made for each field, but it was understood that a single calculation method was used in all PPSE P121 score types, since the same B and the same X and S values were obtained in all fields.

The X, S and B values in the formula used in PPSE P121 and P10 score calculation are not published by MSPC. These values were found in the following manner: Since the formula is a

first order equation, there is a linear relationship between WSS and P121. A line can be created by using the data of 2 exam results related to the exam for a specific year. When the slope of this line and the point where it intersects the y axis is known, P121 corresponding to each WSS can be calculated. Based on the correct and incorrect number of items in the exam result, the candidate's WSS can be calculated but the error in the estimation of the correct items may be large due to the test mean and standard deviations and the rounding made by MSPC in the 5th digit after the comma in the calculated P121. In order to minimize this error, the error was reduced to 3 out of ten thousand by using 70 exam results for each exam year.

The following method was followed to determine how many people were appointed in the January 2020 Contractual Teacher Appointments with the results of 2018 and 2019 exam scores and the exam scores that provided the basis for these appointments. The field-based and institution-based base scores lists, created after these appointments and published by the Ministry of Education, were used to determine the rank of the last appointed person from a specific teaching field with 2019 scores. It was assumed that before the last candidate assigned with 2019 points, all candidates from a field were assigned with 2019 scores as well. It can be argued that this number was an upper limit for 2019 and the actual number would actually be lower because some candidates did not select a post, took both 2018 and 2019 exams or were actually assigned with a 2018 score. For example, the highest-ranking candidate in 2019 exam who scored 100 points was regarded to be appointed in this calculation, although he/she was not included in the appointment lists.

The lowest field-based and institution-based scores obtained as a result of appointments made by the Ministry of Education using 2018 and 2019 exam scores were accessed on the Ministry's website. The consent of the candidates whose PPSE results and scores were used in the study was obtained to be used in this study.

3. RESULT / FINDINGS

To answer the first research question, the parameters used in P-121 calculation were estimated with the operations described in the method section. Table 1 presents the obtained values.

**Table 1.** *Parameters Used in 2018, 2019 and 2020 PPSE P121 Calculation.*

Year	B	$X + 0.5 \cdot S$
2018	81.857283	54.987993
2019	89.525328	55.070568
2020	83.029900	54.812020

The calculation was used to find the slope of the line produced by the P121 formula and the point where this line intersected the y-axis. Since the formula aimed to increase the score of the candidate with the highest WSS to 100, thanks to this relation, the WSS of the person who scored 100 could be found. Again, with the formula, the score of the candidate with a WSS over 0.5 standard deviation was found to be 70. Although X and S values could not be found separately, it was possible to calculate $X + 0.5 \cdot S$ value in this way.

Among the parameters used in the formula, X, which is the average of WSS, and S, its standard deviation, can be expected to be 50.0 and 10.0, respectively, since they are the values obtained from T scores. However, since candidates other than teacher candidates also participated in GC and GA tests, it was found that they differed from these values with small deviations. In this formula, B value was the most important parameter that would affect P121 in response to a WSS. The drawn line passed the point of score 70 around 55 WSS in all the 3 years, but there was a significant difference between years in the B value, which was the main value that would

determine the slope. While this value was 81.857283 in 2018, it increased to 89.525328 in 2019 and dropped to 83.0299 in 2020. In the form of a chart, Figure 1 presents the differences in score caused by these differences. The chart includes weighted standard scores on the horizontal axis and P121 scores corresponding to these WSS scores on the vertical axis.

**Figure 1.** *The Change in P121 Scores Received by the Same WSS Values by Year.*
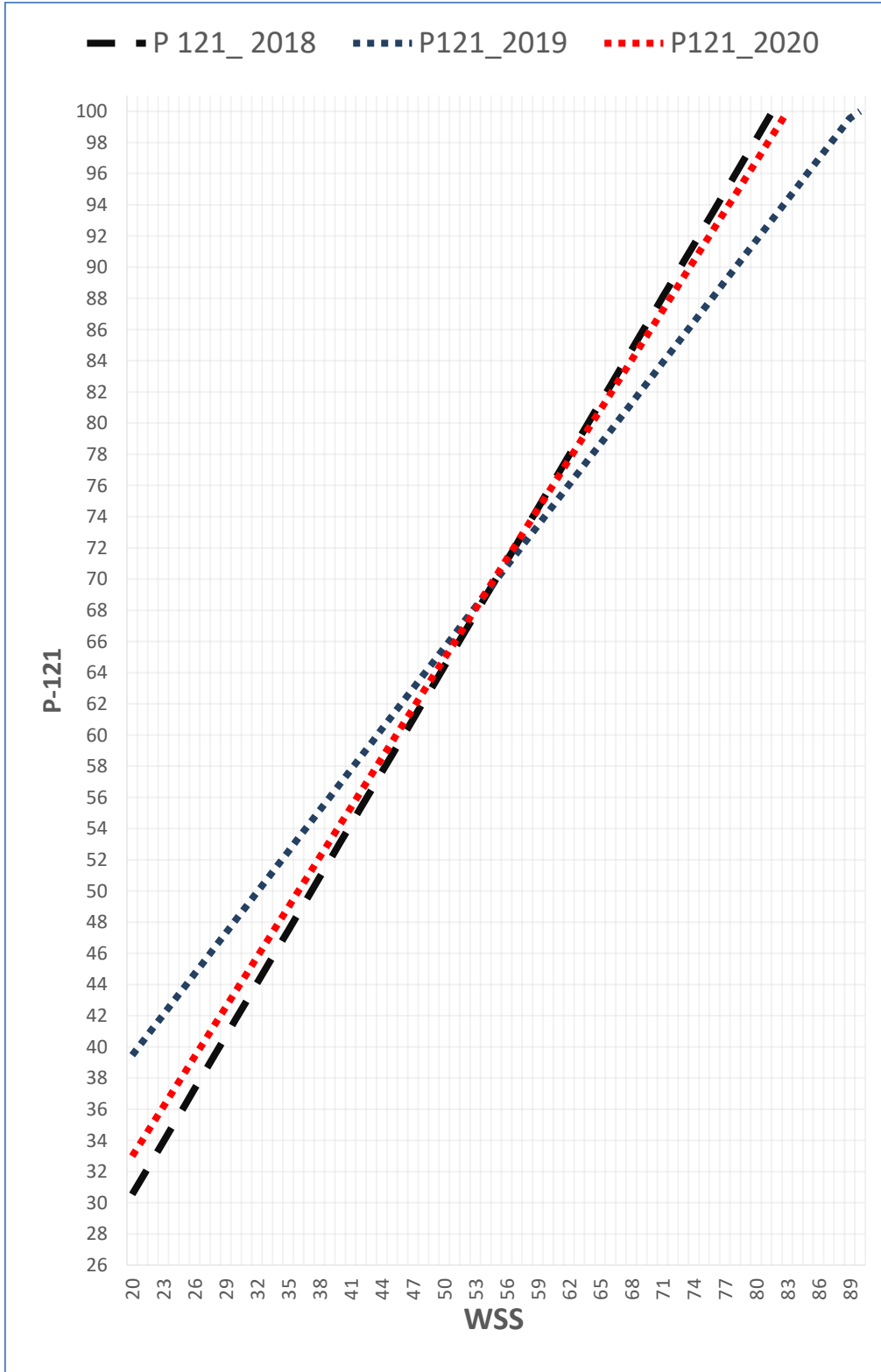
The answer to the second research problem can be seen in Figure 1, B value affected the slope of the line. High B value generated lower P 121 values for the WSS which were bigger than half standard deviation of the mean and generated higher P121 values for smaller WSS. In other words, those with a P121 score above 70 received lower scores than they should have while those below 70 scored higher. This situation was reversed when the B value was low. In that case, those who scored over 70 had high scores and those below 70 had low scores. If a candidate performed very high in any of these 3 years and achieved a very high WSS, other individuals with high WSS scores in the same year would receive lower P121 scores than they would get in another year's exam. This is beyond standardizing individuals' scores using mean and standard deviation. Mean and standard deviation values were obtained from the scores of all candidates taking the exam. While the number of candidates was approximately 400,000, the score of a single person with the highest score in the P121 calculation affected all other candidates' scores. The fact that B value affected the scores that much caused a significant difference on the 2019 exam as can be seen in Figure 1. For example, the P121 of the candidates with a WSS of 80, that is, approximately 3 standard deviations above the candidates who took the exam with them, were 98.20978 in 2018, 91.70623 in 2019 and 96.77874 in 2020. If these three candidates selected teaching posts in the same appointment period, the 2019 candidate would be far behind the others since the appointments were score based and would even fall behind the candidates with 75 WSS in 2018 and 76 in 2020, despite having a higher level of achievement.

The situation was not different when the same chart was prepared according to the Z scores, which expressed in standard deviation the distance of the candidates to the WSS average so that the changes that may occur due to the differences of X and S values of each examination year could be included in the calculation (Figure 2). The horizontal axis included the Z scores calculated using the WSS received by the candidates and the mean and standard deviation of the whole WSS distribution showing their position within the WSS distribution while the vertical axis included the P121 scores corresponding to these Z scores. Analysis of the chart shows that the most disadvantaged group was the candidates of 2019, with scores above 70, the score that set the basis for higher number of appointments. An individual who performed 2.5 standard deviations above the average could obtain a score of 92.50245 in 2018 and a score of 90.46371 in 2020, while the equivalent of the same performance was a score of 87.65991 in 2019.

Apart from the effect described above, there is another aspect of the effect of B value on scores that causes unfairness. All candidates answered 4 tests. 3 of these tests were common, but TCKT tests were different tests consisting of different questions for each teaching field. For example, 18 different TCKT score types were defined in the 2019 PPSE (Measurement, Selection and Placement Center, 2019). The means and standard deviations of these tests were naturally different from each other. Despite the fact that these 18 score types were defined in the 2019 guide, unlike in 2018, and it was stated in the same guide that "PPSE score distribution will be obtained for each PPSE score type out of 100", it was understood that a single B was used in the calculations for all teaching fields. Therefore, this B value belongs to only one of the fields. The scores of the candidates in the other 17 teaching fields were determined according to the performance of the person with the highest score in a test they may or may not have taken. This situation is unacceptable in terms of principles in assessment and measurement. Table 2 displays the Z and T scores that the candidates would get from the TCKT test in 2019 if all questions were answered correctly.

Figure 2. *Change in P121 Scores Obtained by Candidates according to Distance to the WSS Average by Year.*
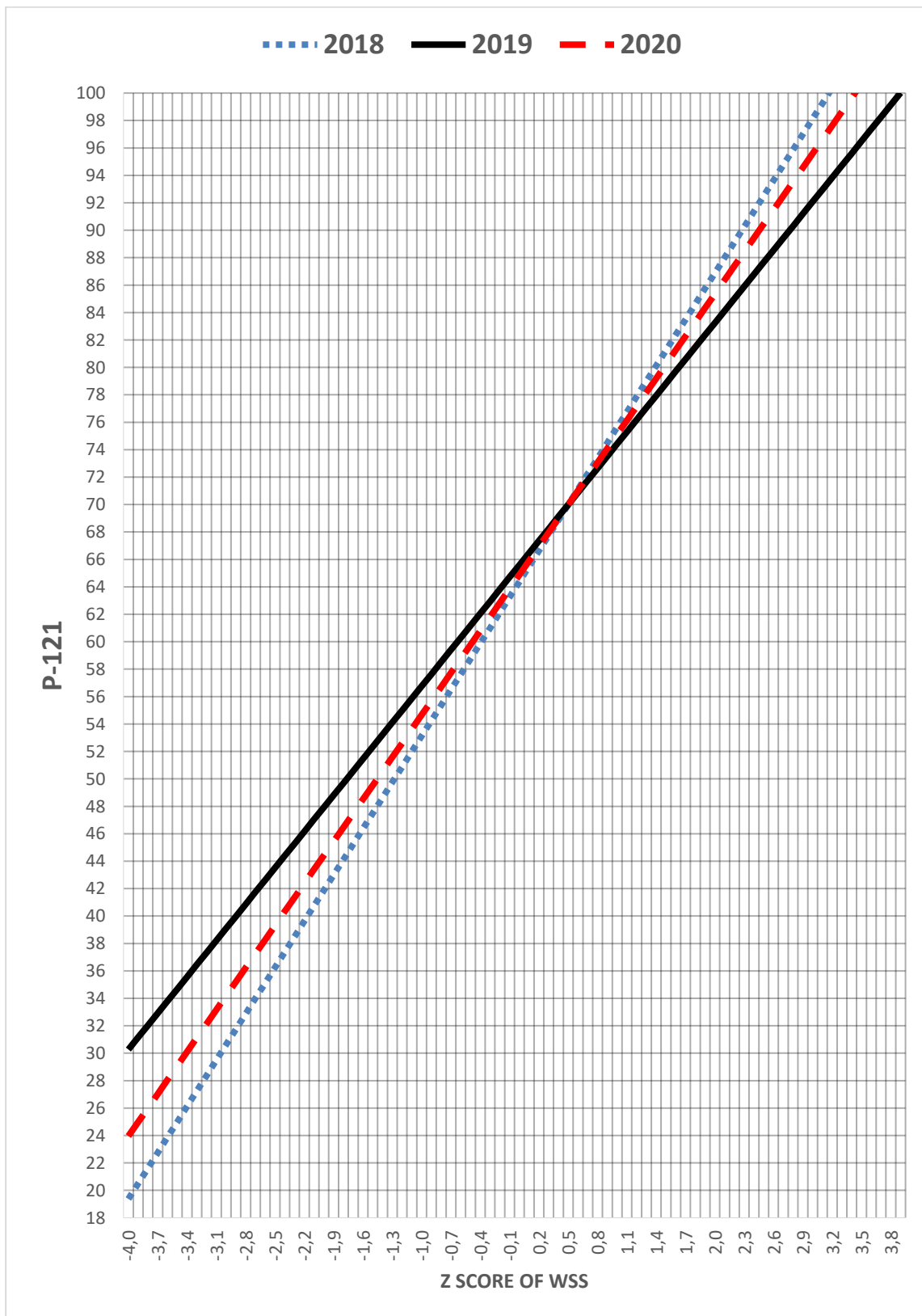
Table 2. *Z and T Scores Corresponding to 75 Correct Answers in TCKT Tests in 2019.*

Tests	Mean	S	Score that can be obtained with 75 correct answers	
			Z	T
Turkish Language Teaching	48.424	10.196	2.606512	76.06512
Elemantry Mathematics Education	30.693	10.425	4.250072	92.50072
Science / Science and Technology	24.496	8.373	6.031769	110.3177
Social Studies	37.551	11.579	3.234217	82.34217
Turkish Language and Literature	27.951	13.867	3.392875	83.92875
History	32.482	13.628	3.119900	81.19900
Geography	35.347	12.424	3.191645	81.91645
Mathematics (High School)	24.268	12.425	4.083058	90.83058
Physics	32.032	15.711	2.734899	77.34899
Chemistry	28.278	14.364	3.252715	82.52715
Biology	25.229	10.802	4.607573	96.07573
Religious Culture and Ethics	42.599	11.592	2.795117	77.95117
Foreign Language (English)	33.863	14.283	2.880137	78.80137
Counselor	50.568	11.481	2.128038	71.28038
Classroom Teaching	32.434	8.823	4.824436	98.24436
Pre-school Teaching	37.382	10.578	3.556249	85.56249
Physical Education Teaching	26.514	8.256	5.872820	108.7282
Religious Vocational School Vocational Classes	38.082	10.353	3.565923	85.65923

As Table 2 shows, the arithmetic means of TCKT tests in the 2019 PPSE varied between 24,268 and 50,568. Candidates who achieved high success in difficult tests could get very high Z and T scores because they were far away from the average of the distribution. In easy tests, on the other hand, Z and T were lower because there was not much difference between the average and 75, the number of questions in the test. For example, a candidate taking the counseling teacher TCKT can have a score of 71.28038 T with 75 corrected score, while the candidate who answers all questions correctly in the science TCKT test can get 110.3177 points. The difference in B will cause unfair decisions from year to year even if the candidates take the same tests. Although it is clear enough, some people would still argue that "if someone can perform very high in an exam, the other candidates after him/her would of course score lower". There are two arguments against this supposition:

1. While this high B value decreases the scores up to 70, it increases the scores below 70, 2. It is not likely to get high scores in some tests anyway. It is clear that calculating the scores of the individuals with a parameter of a test whose average and standard deviation are different from the test they have taken (since candidates can only take one TCKT and no chance of taking the test whose results are being used to calculate results) and comparing these scores with the scores from another year that are valid for 2 years can lead to serious unfairness among candidates. Even if the candidate answered all the questions in the GC, GA, ES and TCKT tests correctly in 2019, a candidate could not have exceeded the B value of 89.525328, used in the score calculation of this exam in 14 of the 18 fields where teachers were appointed with this score type. Since other tests were common, it was necessary to have a high net in a TCKT area with a low average and standard deviation in order to be the first in the exam, others would not have such a chance.

The above section discussed the reasons for the unfair scoring in the exams held in different years due to the calculation method in the P121 formula (research problem 3). Below, the

consequences of these inequalities were presented through practical examples. Table 3 presents the field-based base score and number of appointments for some teaching fields in January 2020 (Ministry of National Education, 2020). The table also includes the WSS required to get a base score in the exams of the last 3 years.

Table 3. *Field Based Base Scores used in January 2020 Contractual Teacher Appointments and the Required WSS to Get These Scores**

Field	Number of Appointments	Base Oral Exam Score	Base PPSE Score	Required WSS to Get the Base Scores		
				2018	2019	2020
Biology	93	81	80.54589	64.33843	**67.18244**	64.73144
Geography	147	81	81.03726	64.77410	**67.74677**	65.19362
Science	1002	76	76,01989	60.32549	**61.98436**	60.47430
Physics	156	75	74.96992	59.39454	**60.77848**	59.48671
Primary School Mathematics	1699	74	74.02994	58.56111	**59.69892**	58.60257
English	1731	71	70.82538	55.71981	**56.01851**	55.58837
Chemistry/ Chemical Technology	151	77	76.54140	60.78788	**62.58331**	60.96483
Mathematics (High School)	501	79	79.26013	63.19842	**65.70575**	63.52206
Pre-school	1513	77	77.44372	61.58792	**63.61962**	61.81355
Counseling	1257	79	79.05386	63,01553	**65,46885**	63,32804
Classroom Teaching	3007	75	74.55471	59.02639	**60.30162**	59.09616
Social Studies	684	80	79.68232	63.57275	**66.19064**	63.91917
History	197	81	81.16877	64.89070	**67.89781**	65.31732
Turkish Language and Literature	384	83	82.92439	66.44731	**69.91413**	66.96865
Turkish Language Teaching	1293	77	77.22951	61.39799	**63.37360**	61.61207

*The fields for which there are no TCKT Exam or where the Foreign Language Exam test score is used while calculating the P121 are not included. Physical Education and Religious Culture and Ethics, whose calculation of WSS-P-121 score conversion can be predicted with relatively more errors due to cancelation of items, are also not included in the table.

Examination of the scores in Table 3 points to the need for a higher WSS in 2019 in order to have the score of the candidate who was appointed the last in all fields. Since this difference was caused by the B value in the calculation and the base scores were all higher than 70, 2019 was the most disadvantaged year in which the largest value B was used. This disadvantage was the greatest in the field of Turkish language and literature, where the score required for appointment was the highest. For candidates who took the 2019 exam to be appointed, their performance in the exam should have been 3.47 standard scores higher than that of 2018 and 2.95 standard scores higher than that of 2020. This difference was lower in the field of English language teaching, where base score for appointment was close to 70. There is another point to take into consideration here: Although the most important decision in teacher appointments is related to whether teachers would or would not be appointed in the first place, there is also the issue about where they would be appointed. Since this decision is also taken according to the level of their scores, the appointed candidates may not be assigned to their first choices if they do not have a very high score.

Table 4 presents the contractual teacher appointments of June 2020 for some teaching fields with the base score and the number of appointments together with the WSS required to get the base score in the exams of the last 3 years (Ministry of National Education, 2020).

Table 4. *Field Based Base Scores used in June 2020 Contractual Teacher Appointments and the Required WSS to Get These Scores* *

Field	Number of Appointments	Base Oral Exam Score	Base PPSE Score	Required WSS to Get the Base Scores		
				2018	2019	2020
Biology	94	80	79.81466	63.60090	**66.34263**	64.04365
Geography	138	80	80.25618	64.08156	**66.84971**	64.45894
Science	1026	74	74.16045	58.67683	**59.84881**	58.72532
Physics	151	74	74.18967	58.70273	**59.88237**	58.75281
Primary School Mathematics	1701	71	70.51645	55.44590	**55.66371**	55.29779
English	1739	68	67.90726	**53.13248**	52.66707	52.84360
Chemistry/ Chemical Technology	154	76	75.64861	59.99629	**61.55795**	60.12508
Mathematics (High School)	498	78	78.10515	62.17437	**64.37969**	62.43569
Pre-school	1518	76	75.97089	60.28204	**61.92809**	60.42822
Counseling	1373	78	77.87273	61.96829	**64.11234**	62.21708
Classroom Teaching	2831	72	71.76993	56.55729	**57.10332**	56.47681
Social Studies	665	79	78.65655	62.66326	**65.01255**	62.95434
History	201	81	80.56881	64.35875	**67.20876**	64.75300
Turkish Language and Literature	344	82	82.17302	65.78112	**69.05118**	66.26191
Turkish Language Teaching	1300	76	75.99971	60.30759	**61.96119**	60.45532

*The fields for which there are no TCKT Exam or where the Foreign Language Exam test score is used while calculating the P121 are not included. Physical Education and Religious Culture and Ethics, whose calculation of WSS-P-121 score conversion can be predicted with relatively more errors due to cancelation of items, are also not included in the table.

Table 4 shows that all fields of teaching except English were disadvantageous in 2019 in terms of the standard score required to get the base score in that field and to be assigned to a teaching post. This disadvantage was more prominent in fields such as Turkish language and literature, history, geography, biology, where higher scores were required to be appointed. Since the base score in the English field fell below 70 points, the critical threshold produced by the P121 score calculation formula, it worked the other way and candidates with 2018 scores in this field were disadvantaged during this appointment period.

Since the validity of the exam scores is 2 years, candidates who took the 2018 and 2020 exam with a relatively equivalent B value will not make a choice in the same appointment period. However, 2019 candidates will continue to experience the victimization they have experienced with the 2018 candidates in the future appointment periods in which they will apply with the 2020 candidates. In these appointments, 2020 candidates with lower standard scores will be ahead of the 2019 candidates and will be appointed before them.

Below are the scores of some candidates along with their exam results and the scores they would receive in other years' exams if they performed similarly. The aim here is to reveal the unfairness described above with concrete examples.

The social studies teacher candidate, whose score reports given in Figure 3, obtained a score of 78.64483 and could not be appointed to a teaching post in January and June 2020 appointment periods with this score. This candidate's WSS was 64.99914554. If the candidate succeeded in obtaining this standard score in 2018 and 2020, the candidate's scores would be 81.291107 and 80.8305, respectively.

In the June appointment period, the base score remained at 78.65655. The candidate who received the same WSS in 2018 was assigned a teaching post, but this candidate could not be appointed. While this candidate fell behind the candidates who received lower scores in 2018 in the appointment periods, he/she will still be behind the candidates with lower standard scores in the 2020 exam in the appointment periods for 2021, and probably will not be appointed to any teaching post with this score.

The history teacher candidate whose score report is presented in Figure 4 had a P121 score of 80.17244. The candidate could not be appointed because he/she fell approximately 0.4 scores behind the last appointed person in the June 2019 appointment period. The candidate's WSS was 66.75334794. If the candidate with the highest WSS in this exam from another TCKT field had received the score that could be received in the 2018 exam or in the 2020 exam, the P121 that the history teacher candidate would get with this WSS would be 83.26955 and 82.695549, respectively. Then, the candidate would be above the base score with both of these two scores, but he/she fell behind those with lower standard scores and could not be appointed.

The Turkish Language teacher candidate with the above score report (Figure 5) could not be appointed to a teaching post due to obtaining only 0.003 points lower than the base score in June 2020 appointment period. However, the equivalent of this performance was WSS 61.95714931. Despite having a WSS score of about 1.2 standard deviations higher than those who took the exam with him/her, the candidate was even behind the 60.30759 WSS candidates in the 2018 exam. If he/she had obtained the same WSS in the 2018 exam, the candidate would have been appointed as a teacher with a score of 77.86016.

The score report presented in Figure 6 is from the field of biology in 2019. The candidate scored 79.80467 and ranked 84th among the candidates who took the exam in this field but was not appointed to a teaching post with 187 open positions during the 2 appointments periods in 2020. The fact that the candidates with 2018 exam scores were placed ahead of the candidate who ranked 84th in the 2019 exam after participating in 2 previous appointment periods can only be explained by the unfair scoring between the exams held in these two years. Since the B value in the 2018 exam was lower than that of the 2019 exam, the 2018 exam candidates who received a lower standard score were placed ahead of this candidate's score. However, if the same candidate had received the same standard score in the 2018 exam, his P121 score would be 82.79361 and would be within the limits that would make it possible for him/her to be appointed. The candidate could only be appointed to a position he/she did not choose in the additional appointment periods available in 2020.

It is possible to present hundreds of examples in this regard. This outcome was experienced because the PPSE of 2019 had an extremely high B score when compared to other years and this B value in the score calculation had no function in formula other than ensuring that the highest score was 100.

Figure 3. *A Score report from Social Studies Field in 2019.*

Education Level	Bachelor's Degree
Date of Exam	14th, 20th, 21st, 28th July 2019

NUMBER OF TRUE AND FALSE ANSWERS FOR EACH TEST									
General Aptitude		General Culture		Educational Science		Teaching Field Knowledge (Morning)		Teaching Field Knowledge (Noon)	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
28	8	47	12	57	22	61	12	-	-

Public Administration		International Relations		Labor Econ. & Industrial Rel.		Law	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-

Economy		Finance		Business		Accounting		Statistics	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-	-	-

SCORES AND RANKS											
SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates
P1	73.03193	112103	602945	P2	75.12638	78923	602945	P3	77.39235	54883	602945
P10	78.44819	33551	356471	P121-4	78.64483	867	17547				

Figure 4. *A Score report from History Field in 2019.*

Education Level	Bachelor's Degree
Date of Exam	14th, 20th, 21st, 28th July 2019

NUMBER OF TRUE AND FALSE ANSWERS FOR EACH TEST									
General Aptitude		General Culture		Educational Science		Teaching Field Knowledge (Morning)		Teaching Field Knowledge (Noon)	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
29	6	52	6	66	11	56	17	-	-

Public Administration		International Relations		Labor Econ. & Industrial Rel.		Law	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-

Economy		Finance		Business		Accounting		Statistics	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-	-	-

SCORES AND RANKS											
SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates
P1	77.94967	42027	602945	P2	80.39786	23624	602945	P3	83.07494	13059	602945
P10	86.70711	3730	356471	P121-4	80.17244	269	19936				

Figure 5. *A Score report from Turkish Language Teaching Field in 2019.*

Education Level	Bachelor's Degree
Date of Exam	14th, 20th, 21st, 28th July 2019

NUMBER OF TRUE AND FALSE ANSWERS FOR EACH TEST									
General Aptitude		General Culture		Educational Science		Teaching Field Knowledge (Morning)		Teaching Field Knowledge (Noon)	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
27	4	43	9	55	19	63	10	-	-

Public Administration		International Relations		Labor Econ. & Industrial Rel.		Law	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-

Economy		Finance		Business		Accounting		Statistics	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-	-	-

SCORES AND RANKS											
SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates
P1	74.38420	87622	602945	P2	75.93372	67280	602945	P3	77.59316	52624	602945
P10	78.13399	35699	356471	P121-4	75.99621	2251	16481				

Figure 6. *A Score report from Biology Field in 2019.*

Education Level	Bachelor's Degree
Date of Exam	14th, 20th, 21st, 28th July 2019

NUMBER OF TRUE AND FALSE ANSWERS FOR EACH TEST									
General Aptitude		General Culture		Educational Science		Teaching Field Knowledge (Morning)		Teaching Field Knowledge (Noon)	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
47	6	40	13	52	26	49	24	-	-

Public Administration		International Relations		Labor Econ. & Industrial Rel.		Law	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-

Economy		Finance		Business		Accounting		Statistics	
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
-	-	-	-	-	-	-	-	-	-

SCORES AND RANKS											
SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates	SCORE NAME	SCORE	RANK	No of Candidates
P1	87.52383	2452	602945	P2	87.29447	2757	602945	P3	87.00006	3728	602945
P10	82.99804	11601	356471	P121-4	79.80467	84	5568				

Another investigation performed in the framework of this study included the comparison of 2018 and 2019 exam scores with the number of appointed teachers (research problem 4). However, since the necessary data for this analysis were not published, a specific method was followed which was mentioned in regard to the limitations in the method section. The number that was attained as the number of candidates appointed with 2019 exam scores was only an upper limit, the actual result is likely to be lower than that figure. For some fields, the calculation could not be completed because the candidate last assigned to a teaching post could not be reached. Table 5 displays the data on the fields for which calculations can be performed.

Table 5. *2019 and 2020 Base Scores and Appointment Rates in Some of the Teaching Fields.*

Field	Total Quota for 2020 Appointments	2019 Base Score	2020 Base Score	Appointed with 2018 Score		Appointed with 2019 Score	
				Number	%	Number	%
Biology	187	81.57524	79.81466	104	55.62	83	44.38
Geography	285	81.62987	80.25618	94	32.92	191	67.08
Science	2028	75.22787	74.16045	141	6.95	1887	93.05
Physics	307	75.45017	74.18967	72	23.45	235	76.55
Chemistry/ Chemical Technology.	305	77.88505	75.64861	107	35.08	198	64.92
Mathematics (High School)	999	80.89463	78.10515	328	32.83	671	67.17
Counseling	2630	79.42048	77.87273	400	15.21	2230	84.79
Social Studies	1349	80.28933	78.65655	486	36.03	863	63.97
History	398	81.48067	80.56881	194	48.74	204	51.26
Turkish Language and Literature	728	83.60853	82.17302	392	53.85	336	46.15
Turkish Language Teaching	2593	77.17681	75.99971	345	13.31	2248	86.69

Table 5 shows that the base scores decreased in all teaching fields included in the table for 2020 appointments. Except for the field of teaching English, which is the only area not included in this table and previously determined to be more advantageous in 2019 exam scores, all base scores dropped in 2020 appointments. For instance, if we take the field of biology teaching as an example, no candidates who participated in the appointment periods with their 2018 exam scores above 81.57524 points should have been left for future periods, (if they had selected their preferred teaching positions and if they had selected enough number of positions to be appointed). In 2020 appointment periods, the candidates with 2019 exam scores which were higher than this base score should have been assigned to posts first, and if there was still a quota, the two groups should have been assigned to these quotas in a mixed manner. However, since the B value of the highest scoring WSS in the 2019 exam, was approximately 7.7 points higher than that of 2018, the scores of the applicants from this field were not very high. Even if all questions were answered correctly in some field tests, it was not possible to score as high as one could in the field of science, after all. After a small number of candidates with 2019 points were appointed, the candidates with both years' scores were placed in their preferred posts and the quota was filled with candidates who scored 79.81466. The ratio of appointments shows

that 44.38% of those appointed from this field were the candidates from 2019 the most. The percentage of 2018 candidates already included in two prior appointment periods and appointed with 2019 candidates was found to be much higher than expected.

It is possible to explain the fact that 2019 candidates were appointed relatively in higher numbers in fields such as science, counseling, and Turkish, where both the number of candidates taking the exam and appointment quota was high: Since the number of applicants was 17,460 in science, 16,916 in counseling and 16,548 in Turkish, it can be expected that the number of people who scored above the base score of the previous year would be higher than a field where less candidates were available such as the field of biology with 5662 candidates. Another reason was related to the fact that the base scores in these fields were closer to 70 where the difference between years was zero. In the case of Turkish language and literature where the number of applicants was high and which has the highest base score, more than half of those appointed were 2018 candidates, despite having lower standard scores just was the case in the field of biology. In addition, since the standard deviation of the Turkish language and literature field test was as high as 13,867, candidates who took the exam in this field did not have the opportunity to get a higher standard score even if they answered almost all of the questions correctly.

Since the B value used when calculating the 2020 P121 scores was approximately 6.5 points lower than that of 2019, 2020 candidates will be advantageous in both appointment periods in 2021 in which the 2019 and 2020 candidates will apply and 2019 candidates will not be appointed although they obtained high standard scores.

To answer the research question 5, the 2018 and 2019 exams were also examined in terms of fields where appointments were made with P10 scores in addition to the P121 scores in the TCKT test. P10 is calculated over the weighted standard scores found by converting the scores of general culture, general ability and educational sciences tests into T scores and multiplying them by the coefficients 0.3, 0.3 and 0.4. Since the same formula is used, the scores are affected by the B value, which is the score of the candidate with the highest WSS, as in the P121 calculation.

Table 6. *B Values Used in 2018, 2019 and 2020 PPSE P10 Calculation.*

Year	B
2018	77.142393
2019	80.683712
2020	79.483411

When the B values were examined by years presented in Table 6, it was found that the highest value of B was obtained in the 2019 exam while the lowest value of B was in the 2018 exam. 2018 candidates experienced a very disadvantageous situation over P121 points previously described in a detailed manner in areas with a base score over 70. 2018 candidates were advantageous in 2018-2019 mixed appointments in these fields. 2020 candidates will be more advantageous in 2019-2020 mixed appointments as well. The advantageous group in a small number of fields such as teaching music with a base score below 70 is the 2019 candidates.

Another issue that may create unfairness in the 2019 PPSE was experienced in the fields of physical education teaching, and in religious vocational high school vocational courses. In 2018, there was no TCKT in these fields and the basic score for appointment for a post was the P10 score. In 2019, TCKT became mandatory in these two fields and the base score for appointment for a post changed to P121. It is considered to be a problematic practice to treat

the P10 scores calculated with the standard scores obtained from 3 tests as the equivalent of P121 scores obtained from 4 tests without applying any equalization procedure in the mixed assignments of 2018 and 2019 candidates.

4. DISCUSSION and CONCLUSION

The following results are obtained based on the investigation described in a detailed manner in the findings and interpretations sections of the PPSE P121 and P10 scores which constitute the basis for teachers' appointment to teaching posts and specific teaching institutions.

The teaching fields where appointments were done with the P121 score, unfair decisions were observed in the 2018-2019 mixed appointment process due to the existence of different tests for each field, the differences in the test statistics for these tests and the effect of the weighted standard score of a candidate over the scores of all candidates. The reason for the unfairness observed here is not only because the statistics of the test items, which is the limitation of the classical test theory stated by Embreston and Reise (2000), depend on the sample. The main reason is to make an extra point conversion after choosing the way of "standardizing the scores by freeing them from the effects of the mean and standard deviation" (Tekin, 1996; Turgut & Baykul, 2010) suggested by the Classical Test Theory.

Some TCKTs were difficult for the candidates and the average was low. In some others, the average was higher. Hambleton & Swaminathan's (1985) criticism that individuals' abilities may vary depending on the items they answer and that they may perform differently in tests of different difficulty is a criticism of using the scores of two different tests that are claimed to measure the same thing together. Here, not only are the difficulties of the tests different, but also the features they measure. In high-average tests, even answering all the questions correctly was not enough to get a high standard score. When calculating P121, using the B value obtained in one of the 18 fields to calculate the scores of the other fields have resulted in lower scores in the TCKT fields with a large mean and standard deviation. This causes field-based unfairness while using the scores from different years in the same appointment period.

Since the B value calculated for 2019 was much higher than that of 2018 and 2020, this created an unfairness against the 2019 candidates in fields with a base score over 70. This situation, also noted in Çelen's (2013) study which presented how the differentiation of the B value would cause a problem in comparing the scores obtained in different years, resulted in a very high level of unfair decisions due to the difference of approximately 8 points in the B value.

The inequalities were also reflected in the number of appointments in related fields. Although 2018 candidates in some fields were included in the appointment for the 4th time, they were appointed at a higher rate than 2019 candidates who were only included in the first two appointments.

Similar investigations show that the scores obtained in the 2019 exam with the highest B value were disadvantageous compared to other fields in areas with a base score of more than 70 even when the P10 was used for appointments in the fields. 2018 candidates were also disadvantaged in a small number of fields with base scores lower than 70 such as teaching music. In the mixed appointment period where 2019-2020 candidates will participate, 2020 candidates will be advantageous in high scoring fields.

TCKT was added to the fields of physical education and religious vocational high school vocational courses in 2019 creating another unfair application by matching the 2018 P10 scores obtained from 3 tests with the P121 scores obtained from 4 completely different tests.

The cases examined in this study which believed to create unfairness should not be considered only in relation to being appointed in a specific period or not. It is an undeniable fact that

unfairness in the scores will affect not only being appointed, but also being appointed at an earlier appointment period and being appointed to a higher-ranking institution.

The following suggestions can be made to the Ministry of National Education and MSPC based on the results of this study which investigated the unfair decisions resulting from the PPSE P121 and P10 formulas and from using a single score for all the candidates who took different field knowledge tests. After calculating the weighted standard scores of the candidates, announcing these scores as an exam result without using a conversion formula can eliminate the inequality between years within the limitations of the classical test theory. If this cannot not done, calculating a separate score for each field can mitigate the inequality, even if it does not fully eliminate it in fields that require teaching field knowledge tests.

Using an exam score in the same application with another exam score without using any equating procedure will cause problems in any case. For this reason, the practice that allows for the validity period of the exams to be longer than one year should be abandoned and this practice should never be utilized again. New calculations can be made using the actual exam data held by MSPC and MoNE, the number of candidates who suffered from the errors in the calculation method can be identified and these grievances can be eliminated by giving them the chance to apply for additional appointments. Similar investigations should also be conducted in regards to university entrance exam where the same or similar formulas are used and for exams such as academic personnel and postgraduate education entrance exam (ALES) which has a validity period of 5 years.

ORCID

Ümit Çelen iD https://orcid.org/0000-0001-7014-2221

5. REFERENCES

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich College Publisher.

Çelen, Ü. (2013). KPSS P10 Hesaplama yöntemine ilişkin sorunlar [Calculation method problems of PPSE P10]. *Ankara University, Journal of Faculty of Educational Sciences*, 46 (1), 127-142. https://dergipark.org.tr/tr/download/article-file/508720

Embretson, S. E. & Reise, S. (2000). *Item response theory for psychologists*. Erlbaum Publishers.

Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory principles and applications*. Kluwer.

Kan, A. (2010). Test eşitleme: aynı davranışları ölçen, farklı madde formlarına sahip testlerin istatistiksel eşitliğinin sınanması [Test equation: testing statistical equality of tests measuring same behaviors and having different item forms]. *Journal of Measurement and Evaluation in Education and Psychology,* 1(1), 16-21. https://dergipark.org.tr/tr/download/article-file/65994

Ministry of National Education. (2020). https://personel.meb.gov.tr/meb_iys_dosyalar/2020_03/19115547_taban_puan_yeni.pdf

Measurment, Selection and Placement Center. (1999). *İlk Defa Devlet Memuru Olarak Atanacaklar İçin Seçme Sınavı (DMS) Kılavuzu* [*Selection Examination Guide for First Time Civil Servants*], MSPC. https://www.osym.gov.tr/TR,3564/dms-kilavuzu.html

Measurment, Selection and Placement Center (2019). *Kamu Personeli Seçme Sınavı (KPSS) Kılavuzu A Grubu ve Öğretmenlik* [*Public Personnel Selection Examination (PPSE) Guide Group A and Teaching*]: MSPC. https://dokuman.osym.gov.tr/pdfdokuman/2019/KPSS/kilavuz22072019.pdf

Tekin, H. (1996). *Eğitimde Ölçme ve Değerlendirme* (9. Baskı) [*Measurement and Evaluation in Education*], Yargı Yayınları.

Turgut, M.F., & Baykul, Y. (2010). *Eğitimde Ölçme ve Değerlendirme* [*Measurement and Evaluation in Education*], Pegem Akademi.