Jurnal Pembelajaran Matematika Inovatif
Volume 9, No. 5, September 2026

ISSN 2614-221X (print)
ISSN 2614-2155 (online)

DOI 10.22460/jpmi.v9i5.33625

ANALYSIS OF THE RELATIONSHIP BETWEEN ACADEMIC REPORT SCORES AND ACADEMIC ABILITY TESTS IN UNIVERSITY ENTRANCE SELECTION IN INDONESIA

R. Bambang Aryan Soekisno*[1], Asya Khaula Aziza[2]

[1,2,3] Master's Program in Mathematics Education, Faculty of Teacher Training and Education, Universitas Singaperbangsa Karawang, Karawang, Indonesia
[1]bambang.as@fkip.unsika.ac.id*, [2]2510632050013@student.unsika.ac.id

ARTICLE INFO

Article History
Received Aug 11, 2026
Revised Sept 10, 2026
Accepted Sept 26, 2026

Keywords:
Academic Report Scores;
Academic Ability Test (TKA);
Assessment divergence;
Mathematics assessment;
University admission

ABSTRACT

This study examines the relationship between five-semester Academic Report Scores and Academic Ability Test (TKA) scores and identifies assessment divergence in the 2026 SNBP selection. A correlational design was employed using secondary data from 124 students from three public senior high schools who were admitted to state universities through SNBP. Data were analyzed using Pearson correlation, quartile analysis, and identification of overestimated and underestimated patterns. The findings revealed a weak but statistically significant positive relationship between academic report scores and TKA scores (r = 0.249; p = 0.005), with approximately 6% shared variance. For Mathematics, the relationship was moderate (r = 0.542; p = 0.005). Quartile analysis showed only 37.9% agreement between the two measures, alongside discrepancies in the estimates of student performance based on report scores. These findings indicate a divergence between school-based assessment and standardized assessment. Therefore, academic report card grades should not be regarded as indicators that can be used interchangeably with standardized academic achievement; rather, they should be integrated to enhance the validity and fairness of the admissions process.

Corresponding Author:
R. Bambang Aryan Soekisno,
Universitas Singaperbangsa Karawang, Indonesia
bambang.as@fkip.unsika.ac.id

Penelitian ini mengkaji hubungan antara nilai rapor akademik lima semester dengan skor Tes Kemampuan Akademik (TKA) serta mengidentifikasi adanya divergensi penilaian dalam seleksi SNBP tahun 2026. Studi ini menggunakan desain korelasional dengan memanfaatkan data sekunder dari 124 siswa dari tiga SMA negeri yang diterima di perguruan tinggi negeri melalui jalur SNBP. Data dianalisis menggunakan korelasi Pearson, analisis kuartil, serta identifikasi pola overestimated dan underestimated. Hasil penelitian menunjukkan hubungan positif yang lemah tetapi signifikan antara nilai rapor akademik dan skor TKA (r=0,249; p=0,005), dengan varians bersama sekitar 6%. Pada mata pelajaran Matematika, hubungan tersebut tergolong sedang (r=0,542; p=0,005). Analisis kuartil menunjukkan kesesuaian kedua ukuran hanya 37,9%, disertai perbedaan estimasi nilai rapor terhadap skor TKA. Temuan ini mengindikasikan adanya divergensi antara penilaian berbasis sekolah dan penilaian terstandar.. Temuan ini mengindikasikan adanya divergensi sebagian antara penilaian berbasis sekolah dan penilaian terstandar. Oleh karena itu, nilai rapor akademik sebaiknya tidak dianggap sebagai indikator yang dapat saling menggantikan dengan capaian akademik terstandar namun perlu mengintegrasikannya demi meningkatkan validitas dan keadilan penerimaan.

This is an open access article under the *CC BY-SA* license

How to cite:

Soekisno, R. B. A., & Aziza, A. K. (2026). Analysis of the relationship between academic report scores and academic ability tests in university entrance selection in Indonesia. *JPMI – Jurnal Pembelajaran Matematika Inovatif, 9*(5), 1339-1354.

INTRODCUTION

Admission to Indonesian state universities through the National Selection Based on Achievement (SNBP) relies primarily on students' Academic Report Scores from semesters 1-5, so school-based assessment has become the main gateway to higher education (Irawan, 2026). In this study, Academic Report Scores refer to the teacher-assigned grades accumulated in students' report cards, whereas Academic Ability Test (TKA) scores refer to the results of a standardized, computer-based national test administered under uniform conditions. Starting with the 2026 admission cycle, report scores are no longer used on their own but are accompanied by TKA results, which are intended to check whether school achievement is consistent with standardized academic performance (BSKAP Kemendikdasmen, 2025b; Kamelia, 2025). The TKA was also designed to serve as a national mapping instrument for educational quality (Juriyah et al., 2025). Because both instruments now jointly inform admission decisions, the extent to which they correspond is no longer a technical detail but a question that directly affects the fairness of SNBP, and it is this correspondence that the present study examines.

One reason for introducing the TKA is the concern that report scores have become inflated and compressed. Grade inflation refers to the awarding of grades that are higher than students' demonstrated achievement warrants; because it is relative, it can only be detected when grades are compared with an external measure of the same competencies (Bejar & Blew, 1981). Evidence from the United States shows that high-school grade point averages rose steadily over time while standardized test scores remained stable or declined, a widening gap attributed to grade inflation (Chen & Sanchez, 2024). In Indonesia, teachers' grading has been described as a "hodgepodge" practice that mixes achievement with effort, attendance, and participation, and this practice has been directly linked to grade inflation in secondary schools (Arrafii, 2020). Teachers in Indonesian secondary schools also report relying on non-achievement factors such as attendance when determining final grades (Zulaiha et al., 2020), and official TKA documents note that report scores tend to cluster in a narrow, high range compared with standardized assessments (BSKAP Kemendikdasmen, 2025b). When grades are inflated and clustered, they lose their capacity to distinguish between students of different ability, which provides a concrete mechanism through which report scores and TKA scores may diverge in SNBP.

Beyond inflation, Indonesian evidence indicates that the quality and standards of school-based assessment vary considerably between schools and teachers. An analysis of school examinations in Mathematics found wide variation in item quality and only a partial contribution of school examination results to national examination outcomes (Sulistiawan, 2016). A study of mathematics teachers in Indonesian secondary schools further showed that teacher-developed assessments were predominantly aligned with lower-order thinking skills rather than the higher-order skills targeted by the curriculum (Zana et al., 2024). Evidence from the national examination itself illustrates how strongly scores depend on who controls the testing conditions: when paper-based examinations were replaced by computer-based testing, school-level scores fell by about half a standard deviation, indicating that earlier results had been inflated by irregular practices (Berkhout et al., 2024). Psychometric studies of Indonesian school tests likewise emphasize that instruments which have not undergone adequate validity and reliability evaluation can produce biased and inconsistent measurement (Ernawati et al., 2024; Ocy et al., 2023). Taken together, these Indonesian findings suggest that a report score of the same value may not carry the same meaning across schools, which is why the relationship between report scores and the TKA should be examined both across and within schools.

The two instruments also differ in how they are constructed and in what they are designed to capture. National assessment policy requires that learning outcomes be assessed comprehensively, continuously, objectively, and accountably, covering cognitive, affective, and psychomotor aspects (BSKAP Kemendikdasmen, 2025a). Report scores therefore summarize a broad range of classroom evidence collected over five semesters. The TKA, in contrast, is developed using psychometric procedures such as item response theory, which provide stronger guarantees of measurement consistency and of the ability to distinguish between students of different ability levels (Hayat et al., 2020). Classroom instruments, by comparison, rarely undergo such calibration (Farozin et al., 2022; Harun et al., 2023). This difference does not make one instrument superior to the other, because the TKA captures mainly cognitive performance at a single point in time and cannot represent the contextual learning processes reflected in report scores (Wibowo, 2025). It does, however, imply that a perfect correspondence between the two should not be expected, and that the size of the observed correspondence is itself informative about how far report scores can be treated as interchangeable with standardized results.

International studies provide converging evidence that teacher-assigned grades can be shaped by factors unrelated to achievement. Comparisons of blind and non-blind grading have shown that teachers adjust scores according to students' classroom behavior (Ferman & Fontes, 2022), gender (Contreras, 2024), and social background such as caste (Hanna & Linden, 2012). A scoping review of 37 studies concluded that such grading discrimination is a recurring, cross-national pattern (Zanga & De Gioannis, 2023), and experimental evidence indicates that bias becomes stronger when grading criteria are vague (Quinn, 2020). These findings come from other education systems and cannot by themselves demonstrate that report scores in Indonesia are biased in the same ways. Read alongside the Indonesian evidence on hodgepodge grading and grade inflation presented above, however, they strengthen the expectation that report scores contain variance unrelated to the competencies measured by the TKA, which is precisely the expectation that this study tests with SNBP data.

Research on predictive validity offers further context but leaves an important gap. Internationally, standardized test scores have been found to predict retention in higher education more consistently than school grades, particularly when grading standards vary across schools (Meeter, 2023), and admission tests have shown substantial incremental validity for academic performance beyond high-school grades (Jaehn et al., 2025). In Indonesia, scholastic aptitude test scores predicted students' cumulative academic achievement more consistently than school-based records (Asrijanty, 2014), whereas studies comparing admission pathways have produced mixed results: report scores were not significantly related to university grade point averages in one study (Lukmanulhakim, 2018), while another found no meaningful difference in academic performance between students admitted through report-based and test-based pathways (Yusnadi et al., 2021). What is known, therefore, is that school grades and standardized scores perform differently as predictors of later university outcomes. What remains unknown is how report scores and TKA scores relate to each other at the point of admission, under the dual-assessment policy introduced for the 2026 admission cycle. This concurrent relationship is a logical prerequisite for predictive questions: if the TKA is to validate report scores, it must first be established whether, and for whom, the two instruments agree.

This question is particularly timely. The TKA was first administered in November 2025 and became a mandatory companion to report scores only from the 2026 admission cycle (Kamelia, 2025), so the 2026 SNBP cohort analyzed here is the first that allows the two instruments to be

compared within the new policy. Earlier Indonesian studies compared report scores with national examination results (Sulistiawan, 2016) or with university grade point averages (Asrijanty, 2014; Lukmanulhakim, 2018; Yusnadi et al., 2021), but none examined the relationship between report scores and TKA scores among students actually admitted through SNBP. The present study differs from these studies in three ways: it uses the newly available TKA data, it examines the relationship separately for each school to reveal between-school differences in grading, and it uses a quartile-based classification to identify individual students whose report scores over- or underestimate their standardized performance. Because SNBP outcomes have long-term consequences for access to higher education and the TKA has been positioned as a tool for accountability in the national system (Andina, 2025), early empirical evidence on divergence between the two instruments is needed to inform how they are weighted and interpreted.

Mathematics is examined separately in this study for three reasons. First, Mathematics is one of the compulsory TKA subjects, and the TKA Mathematics test assesses not only knowledge of facts, concepts, and procedures but also the application of mathematical knowledge and reasoning in solving problems (BSKAP Kemendikdasmen, 2025b). Second, Indonesian students have repeatedly shown difficulties with numeracy tasks that require interpreting graphs, tables, or diagrams and applying reasoning to contextual situations (Kristianti & Handayani, 2023; Setianingsih et al., 2022), and numerical ability has been linked to mathematical problem-solving performance (Gunur et al., 2018). Third, classroom mathematics assessment in Indonesian secondary schools has been found to focus largely on lower-order skills (Zana et al., 2024), so Mathematics report scores may reward procedural mastery more than the reasoning emphasized by the TKA. Previous mathematics education studies have described students' numeracy performance on a single type of assessment; they have not examined whether students' standing in classroom-based Mathematics grades corresponds to their standing on a standardized Mathematics test. Examining this correspondence directly addresses that unanswered question and links the selection issue to the practice of mathematics assessment in schools.

Based on the gap described above, this study aims to examine the relationship between Academic Report Scores and TKA scores among students admitted through SNBP 2026. Specifically, it addresses three questions: (1) how strongly overall Academic Report Scores are associated with overall TKA scores, both in the combined sample and within each school; (2) how strongly Mathematics Report Scores are associated with TKA Mathematics scores; and (3) to what extent students' relative positions on the two measures diverge, as reflected in overestimated and underestimated patterns. The study is expected to contribute empirical evidence on assessment divergence in the first year of Indonesia's dual-assessment admission policy. Practically, the findings can inform policymakers in weighting report scores and TKA results within SNBP, encourage schools to review the consistency of their grading standards, and guide mathematics teachers in designing classroom assessments that capture application and reasoning, so that report scores better reflect the competencies assessed nationally.

METHOD

This study employed a quantitative approach with a correlational design, which was used to determine the strength and direction of the association between school-based assessment and standardized assessment without manipulating any variable. The accessible population consisted of all students from three public senior high schools in Bogor City who were admitted to state universities through SNBP 2026 (N = 125 students). Students were included if they (a)

were admitted through SNBP 2026, (b) had complete Academic Report Scores for semesters 1–5 as submitted to the SNBP database, and (c) had official TKA scores for all subjects they took, including Mathematics. Students were excluded if any semester score or TKA subject score was missing, if they did not sit the TKA, or if their records could not be matched across the two data sources. Data completeness was checked by verifying that each student had all five semester scores and all TKA subject scores. After screening, one student from School A was excluded because of incomplete data, and 124 students met the criteria (School A: n = 61; School B: n = 25; School C: n = 38); this purposive selection based on predetermined criteria follows Etikan et al. (2016). A sensitivity analysis indicated that a sample of 124 provides 80% power to detect correlations of $|r| \geq 0.25$ at $\alpha = 0.05$ (two-tailed), that is, effects between small and medium in magnitude (Cohen, 1988); smaller correlations, and correlations within the smaller school subsamples, could not be detected with the same precision.

The variables were calculated as follows. The Overall Academic Report Score was calculated as the mean of students' scores across all subjects recorded in the SNBP database over semesters 1–5, on a 0–100 scale. The Mathematics Report Score was calculated as the mean of students' Mathematics scores across semesters 1–5. The Overall TKA Score was calculated as the mean of the official TKA scores across all TKA subjects taken by each student, while the TKA Mathematics Score represented the official TKA Mathematics score. Semester Mathematics scores were available only in the records of School B, because the records provided by Schools A and C contained only semester averages across all subjects; therefore, the Mathematics analysis was conducted on this subsample (n = 25). Because students may have different subject combinations in their academic records and TKA selections, the overall scores were calculated based on the subjects available for each student. All scores were obtained from school records as secondary data, and student identities were anonymized before analysis.

The relationship between the variables was analyzed using Pearson product–moment correlation at a significance level of 0.05, because the coefficient measures the linear association between two continuous variables (Schober et al., 2018). Before the analysis, normality was checked using the Shapiro–Wilk test, and linearity and outliers were inspected using scatter plots. Correlation coefficients were interpreted following Schober et al. (2018): $|r| < 0.10$ negligible (very weak), 0.10–0.39 weak, 0.40–0.69 moderate, 0.70–0.89 strong, and ≥ 0.90 very strong. Correlations were computed for the combined sample, for each school, and for the Mathematics scores; for the smaller Mathematics subsample, Spearman's rank correlation and an analysis excluding the most extreme score were used as robustness checks. Because a correlation coefficient describes association rather than equivalence, the coefficient of determination (r^2) was also reported to indicate the proportion of shared variance. All analyses were performed using JASP version 0.19.3.

To identify assessment divergence at the individual level, a quartile analysis was conducted on the combined sample. Students were assigned to quartiles separately for overall Academic Report Scores and overall TKA scores, using the 25th, 50th, and 75th percentiles of each distribution as cut points (Q1 = lowest 25%, Q4 = highest 25%); because no scores were tied at the cut points, each quartile contained exactly 31 students. For each student, the difference between the report-score quartile and the TKA quartile was calculated. A student was classified as aligned when the two quartiles were the same, as showing a minor discrepancy when they differed by one quartile, as overestimated when the report-score quartile was at least two quartiles higher than the TKA quartile, and as underestimated when the report-score quartile was at least two quartiles lower than the TKA quartile. A difference of two or more quartiles was used as the threshold because it cannot arise merely from a student lying close to a single

cut point: it implies a shift in relative standing of at least about 25 percentile points, and typically of about half of the distribution. The classification is relative to the present sample and indicates divergence in students' relative positions, not absolute over- or underestimation of ability.

RESULT AND DISCUSSION

Result

The results are presented in three parts: the correlation between overall Academic Report Scores and TKA scores within each school and in the combined sample, the quartile-based classification of divergence, and the correlation between Mathematics Report Scores and TKA Mathematics scores. Correlations are interpreted as associations between the two scores, following the criteria described in the Method section. School-level results are reported first because, as discussed in the introduction, grading standards may differ between schools. Before the correlations were computed, normality was checked with the Shapiro–Wilk test. Academic Report Scores ($W = 0.984$, $p = 0.157$) and TKA scores ($W = 0.993$, $p = 0.823$) were normally distributed in the combined sample, as were both variables within each school (all $p > 0.05$). In the Mathematics analysis, Mathematics Report Scores ($W = 0.927$, $p = 0.073$) and TKA Mathematics scores ($W = 0.949$, $p = 0.235$) were also normally distributed, so Pearson correlation was appropriate for all analyses.

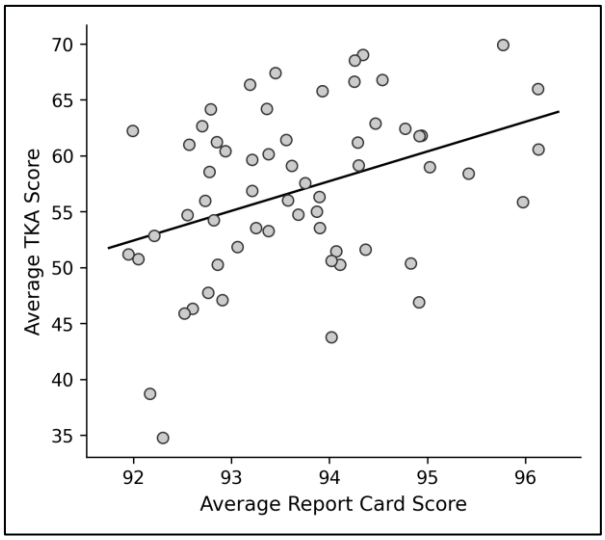


Figure 1. Scatter Plot of the Relationship between Academic Report Scores and Academic Ability Test (TKA) Scores for School A

Figure 1 shows a weak, positive, and significant correlation between Academic Report Scores and TKA scores in School A ($r = 0.377$; $p = 0.003$; $n = 61$). Students with higher report scores in this school tended to have higher TKA scores, but the coefficient of determination ($r^2 \approx 0.14$) indicates that only about 14% of the variance in one score is shared with the other. This result shows that, in School A, students' relative standing on report scores was only partly consistent with their relative standing on the TKA. It should not be read as evidence that report scores in this school accurately represent academic ability, because a correlation describes the association between two scores rather than the validity of either measure, and about 86% of the variance remains unshared. Establishing such validity would require additional evidence, for example on the content of classroom assessments and on their prediction of later academic outcomes.

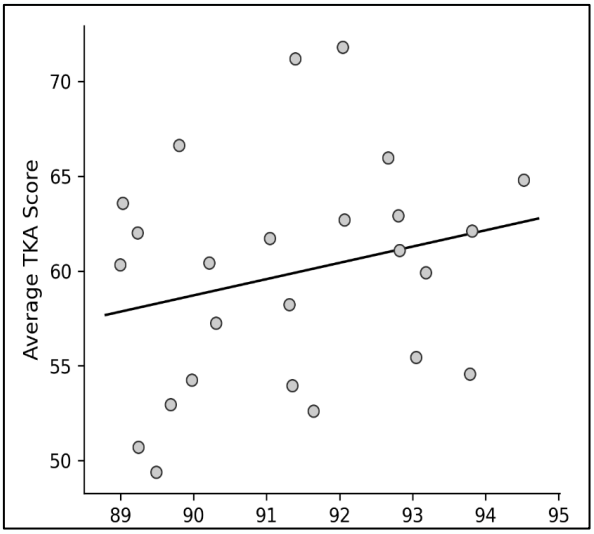


Figure 2. Scatter Plot of the Relationship between Academic Report Scores and Academic Ability Test (TKA) Scores for School B

Figure 2 shows a weak, positive, and non-significant correlation between Academic Report Scores and TKA scores in School B ($r = 0.244$; $p = 0.239$; $n = 25$). The shared variance is about 6% ($r^2 \approx 0.06$), which means that students' report scores in this school provided very little information about their TKA results. Although the direction is positive, the non-significant result indicates that the association cannot be distinguished from zero in this subsample. The smaller size of the school subsample also limits the precision of this estimate, so the result should be interpreted as showing that no clear correspondence was observed rather than as proof that none exists.

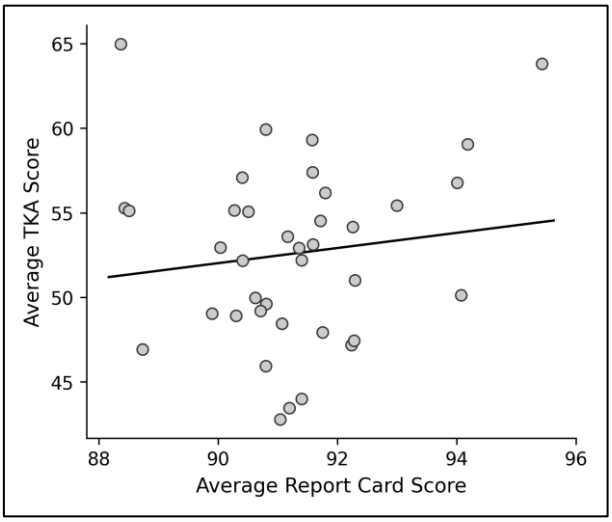


Figure 3. Scatter Plot of the Relationship between Academic Report Scores and Academic Ability Test (TKA) Scores for School C

Figure 3 shows a weak and non-significant correlation between Academic Report Scores and TKA scores in School C ($r = 0.132$; $p = 0.430$; $n = 38$). The coefficient is close to zero, and the shared variance is below 2% ($r^2 \approx 0.02$). In this school, therefore, students' report scores provided almost no information about their TKA scores: students with similar report scores obtained widely different TKA results, and vice versa. Of the three schools, School C shows the weakest correspondence between school-based and standardized assessment.

Comparing the three schools reveals that the association between report scores and TKA scores is not uniform across educational units. Although all three coefficients fall in the weak range, only the association in School A is statistically significant, and its strength declines from School A ($r = 0.377$) to School B ($r = 0.244$) and School C ($r = 0.132$). This variation is consistent with the Indonesian evidence that grading standards, assessment quality, and the weight given to non-achievement factors differ between schools and teachers (Arrafii, 2020; Sulistiawan, 2016). The present data cannot identify which specific practices produced the differences, because information on each school's grading policy was not collected. Nevertheless, the pattern indicates that the degree to which report scores correspond to standardized results depends on the school context, which has direct implications for a selection system that compares report scores across schools.

When the data from the three schools were combined, the correlation between Academic Report Scores and TKA scores was weak but statistically significant ($r = 0.249$; $p = 0.005$), with about 6% shared variance ($r^2 \approx 0.06$). The pooled coefficient reaches significance mainly because the combined sample is larger, not because the association is stronger: it is weaker than the correlation found within School A, a pattern that can occur when schools differ in the general level of their report scores. If one school awards systematically higher report scores than another for comparable TKA performance, combining the schools mixes these different grading scales and weakens the overall association. The scatter plot in Figure 4 illustrates the pooled pattern.

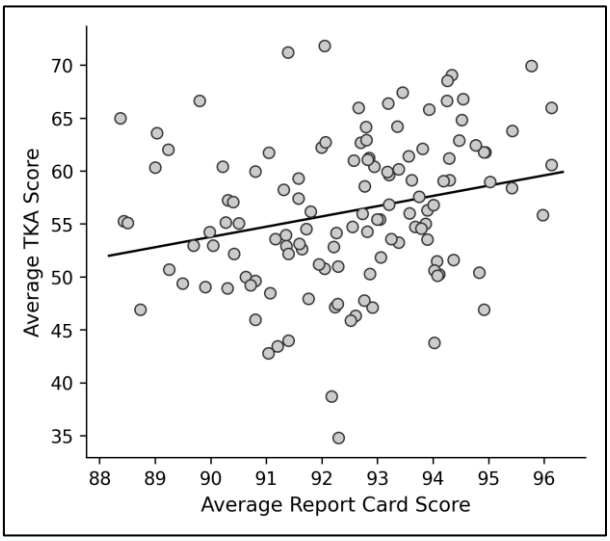


Figure 4. Scatter Plot of the Relationship between Academic Report Scores and Academic Ability Test (TKA) Scores

Figure 4 shows a positive but shallow trend, with the data points widely dispersed around the fitted regression line. Students with similar Academic Report Scores obtained very different TKA scores; several students with high report scores had relatively low TKA scores, while others with relatively low report scores performed well on the TKA. Overall, the figure indicates that, in the combined sample, report scores reflected students' performance on the TKA only to a limited extent, which motivated the quartile analysis of individual divergence presented next.

The quartile analysis compared each student's quartile position on overall Academic Report Scores with the quartile position on overall TKA scores, using the classification criteria described in the Method section. If the two measures ranked students in the same way, most

students would fall on the diagonal of the cross-tabulation, that is, in the same quartile for both measures. Table 1 presents the cross-tabulation.

Table 1. Cross-tabulation of students' quartile positions on overall Academic Report Scores and overall TKA scores (n = 124)

Report-score quartile	TKA Q1	TKA Q2	TKA Q3	TKA Q4	Total
Q4	6	3	9	13	31
Q3	3	9	12	7	31
Q2	10	10	5	6	31
Q1	12	9	5	5	31
Total	31	31	31	31	124

In the upper report-score quartile (Q4), not all students were in the upper TKA quartile. 13 (41.9%) of the students in Q4 for report scores were also in Q4 for the TKA, whereas 9 (29.0%) were in Q2 or Q1 for the TKA and were therefore classified as overestimated. These are students with report scores in the highest-achieving quarter of the sample but TKA performance in the lower half. Across all report-score quartiles, 12 (9.7%) of students were classified as overestimated. This pattern is consistent with the possibility that report scores at the top of the distribution are partly inflated or compressed, so that a high report score does not necessarily signal high standardized performance.

In the middle quartiles (Q2 and Q3), students were spread across all TKA quartiles, and most of the discrepancies in these groups were of one quartile: 31 of the 62 students (50.0%) showed a minor discrepancy, accounting for 31 of the 40 students who were not aligned. Students in the middle of the report-score distribution are the most numerous and the closest to the cut points, so movement between adjacent quartiles is expected even when two measures are related. Nevertheless, the absence of a clear concentration on the diagonal in these groups shows that, for the majority of students, report scores provided little information about their relative position on the TKA.

In the lower report-score quartile (Q1), 10 (32.3%) of students reached Q3 or Q4 on the TKA and were classified as underestimated; across all quartiles, 16 (12.9%) of students fell into this category. These students obtained relatively low report scores compared with their peers but performed above the median on the TKA, suggesting that their classroom grades did not fully reflect the competencies measured by the standardized test. Overall, only 47 (37.9%) of students occupied the same quartile on both measures. The combination of overestimated and underestimated cases at both ends of the distribution illustrates, at the individual level, why a weak correlation leaves considerable room for misclassification, and shows that grouping students by report scores does not reproduce their grouping by TKA performance.

To examine whether the divergence observed at the overall level is also present within a specific subject, Mathematics Report Scores were compared with TKA Mathematics scores. As explained in the introduction, the TKA Mathematics test emphasizes the application of mathematical knowledge and reasoning in solving problems in addition to factual and procedural knowledge (BSKAP Kemendikdasmen, 2025b). This comparison therefore shows whether students' classroom-based standing in Mathematics corresponds to their standing on a test with broader cognitive demands. The results are presented in Figure 5.

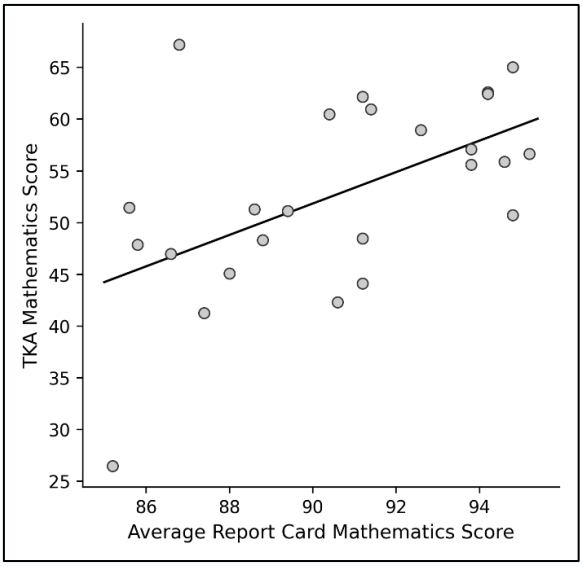


Figure 5. Scatter Plot of the Relationship between Academic Mathematics Report Scores and Mathematics Academic Ability Test (TKA) Scores in School B (n = 25)

The correlation between Mathematics Report Scores and TKA Mathematics scores in School B was moderate, positive, and significant ($r = 0.542$; $p = 0.005$; $n = 25$), with about 29% shared variance ($r^2 \approx 0.29$). The association remained significant when the student with the lowest TKA Mathematics score was excluded ($r = 0.445$; $p = 0.029$; $n = 24$) and when Spearman's rank correlation was used ($\rho = 0.515$; $p = 0.008$). Students with higher Mathematics report scores therefore tended to obtain higher TKA Mathematics scores, and the association was stronger than that between overall report scores and overall TKA scores in the same school ($r = 0.244$). This suggests that a subject-specific report score corresponds more closely to a test of the same subject than an average across all subjects does. However, the difference between the two coefficients was not tested statistically, and about 71% of the variance remained unshared, so Mathematics report scores still did not fully reflect students' standing on the TKA Mathematics test. Because this result is based on a single school, it cannot be generalized to the other schools in the sample.

Discussion

Pada The main finding of this study is that, among students admitted through SNBP 2026, Academic Report Scores showed only a weak association with TKA scores, with the strength of the association varying across schools, whereas Mathematics report scores corresponded more closely with TKA Mathematics scores in the one school where this could be examined. This indicates that overall report scores and TKA scores rank students only partly in the same way and therefore cannot be treated as interchangeable indicators of academic achievement. The finding is in line with Indonesian studies reporting that report scores were not significantly related to university grade point averages (Lukmanulhakim, 2018) and that standardized aptitude scores were more consistent predictors of academic achievement than school records (Asrijanty, 2014). It also parallels international evidence that standardized tests and school grades capture partly distinct aspects of academic readiness (Jaehn et al., 2025). The present study differs from these studies in that it examines the concurrent relationship at the point of admission rather than the prediction of later outcomes, and it shows that the limited correspondence is already visible before students enter university.

One explanation for the weak association lies in differences in grading practices between schools and teachers. Classroom-based assessment depends on the learning context and on teachers' and schools' policies, so its results cannot automatically be equated with standardized tests administered under uniform procedures (Sudjana, 2017). Studies of Indonesian secondary teachers have shown that final grades often combine achievement with effort, attendance, and participation (Arrafii, 2020; Zulaiha et al., 2020). When such practices differ between schools, a report score of 90 may represent different levels of achievement in different schools. The between-school variation found in this study (a significant association in School A but almost none in School C) and the weaker pooled correlation are consistent with this explanation. This interpretation remains tentative because the study did not collect data on each school's grading policies, but it identifies a plausible mechanism that future research can test directly.

A second explanation concerns grade inflation and the restricted range of scores in the sample. When report scores are inflated, they cluster near the top of the scale and lose their ability to distinguish between students (Bejar & Blew, 1981). Chen and Sanchez (2024) documented this process in the United States, where rising high-school grades were accompanied by stable or declining standardized scores, and Indonesian evidence shows that results under school-controlled conditions can overstate achievement, as seen in the drop in national examination scores after the introduction of computer-based testing (Berkhout et al., 2024). In addition, the present sample consisted only of students admitted through SNBP, who are by definition students with high report scores. A narrow range of scores on one variable reduces the observable correlation even when the underlying relationship is stronger in the full population (Sackett & Yang, 2000). The overestimated cases identified in the quartile analysis, namely students with top-quartile report scores but below-median TKA scores, are consistent with this combination of inflation and range restriction. These considerations mean that the correlations reported here describe the relationship among admitted students and may underestimate the relationship among all applicants.

A third explanation is that the two instruments are designed to measure partly different constructs. Measurement validity depends on the extent to which an instrument consistently represents the construct it is intended to measure (Arifin, 2016). Report scores, as authentic assessments accumulated over five semesters, are intended to capture a broad range of competencies, including knowledge, skills, and attitudes, so differences from standardized test results are to be expected (Ani, 2014). The TKA, by contrast, measures cognitive performance on calibrated items at a single point in time (Hayat et al., 2020). From this perspective, the weak association does not imply that either instrument is wrong; it shows that they provide different kinds of information. The problem arises only when one is used as if it were a substitute for the other, as when report scores are treated as a sufficient indicator of the standardized academic ability required for university study.

The Mathematics result suggests that correspondence is higher when the school-based and standardized measures refer to the same subject, but it also shows the limits of that correspondence. The TKA Mathematics test emphasizes the application of mathematical knowledge and reasoning in solving problems (BSKAP Kemendikdasmen, 2025b), whereas classroom mathematics assessments in Indonesian secondary schools have been found to target mainly lower-order skills (Zana et al., 2024). This difference in cognitive demand may explain why about 71% of the variance remained unshared even within the same subject: Mathematics report scores may partly reflect success on routine and procedural tasks, whereas TKA Mathematics scores depend more on the ability to interpret information and reason in unfamiliar contexts. This interpretation is consistent with studies showing that Indonesian students

struggle with numeracy items that require interpreting graphs, tables, and diagrams and translating contextual problems into mathematical reasoning (Kristianti & Handayani, 2023; Setianingsih et al., 2022), and with evidence that numerical ability is closely related to mathematical problem solving (Gunur et al., 2018). Because this result is based on 25 students from one school, it should be confirmed in other schools before conclusions are drawn about Mathematics assessment more generally.

These findings have direct implications for mathematics teaching and assessment. If classroom assessment rewards mainly procedural accuracy, students can obtain high Mathematics report scores without developing the reasoning competencies assessed nationally. Instructional approaches that explicitly target application and reasoning, such as a flipped classroom integrated with discovery learning, have been shown to improve senior high school students' numeracy (Saniah & Nindiasari, 2023). Mathematics teachers can therefore strengthen the correspondence between report scores and standardized results by including contextual and non-routine tasks in their assessments, using clear scoring criteria that separate achievement from effort and attendance, and regularly comparing the cognitive level of their test items with the competencies described in the TKA framework. Such changes would make Mathematics report scores more informative about students' mathematical competence without reducing assessment to test preparation.

For the selection system, the findings indicate a risk of misclassification when admission decisions rely heavily on a single type of evidence. The overestimated and underestimated cases show that some students are ranked considerably higher, and others considerably lower, by report scores than by the TKA. Research on admission instruments indicates that combining school grades and standardized scores yields stronger prediction than either measure alone (Chen & Sanchez, 2024; Jaehn et al., 2025), and that even within a single test, different components vary in their predictive strength (Ferdian & Azwar, 2019). For SNBP, this suggests that report scores and TKA results should be treated as complementary sources of evidence, and that the between-school differences found here could be addressed through procedures that calibrate school grading against standardized results, such as monitoring the agreement between each school's report scores and its students' TKA performance. Strengthening teachers' assessment literacy would also help reduce the influence of non-achievement factors on grades (Zulaiha et al., 2020).

This study has several limitations. First, the data came from 124 students in three schools, so the school-level estimates are based on small subsamples and the findings cannot be generalized to all regions or school types; in addition, the Mathematics analysis was limited to the 25 students of School B, the only school whose records contained semester Mathematics scores. Second, the sample included only students admitted through SNBP, which restricts the range of report scores and may lower the observed correlations. Third, the quartile classification is relative to this sample and depends on the chosen threshold. Fourth, the study did not collect information on schools' grading policies or on students' subsequent academic performance, so the proposed explanations could not be tested directly and predictive validity could not be examined. Future studies should therefore use larger samples from several regions that include both admitted and non-admitted applicants, apply multilevel models that separate within-school and between-school relationships, follow students into their first year of university to examine predictive validity, extend the analysis to other TKA subjects, and combine quantitative analysis with qualitative investigation of teachers' grading practices.

CONCLUSION

This study concludes that, among students admitted through SNBP 2026, Academic Report Scores and TKA scores correspond only weakly and inconsistently, which means that the two instruments cannot be treated as interchangeable measures of academic achievement. The degree of correspondence depends on the school, indicating that report scores do not carry the same meaning across educational units. At the individual level, a substantial number of students are ranked considerably higher or lower by their report scores than by their standardized performance, showing that assessment divergence affects individual admission decisions and not only aggregate statistics. In Mathematics, report scores corresponded more closely with TKA results in the one school where this could be examined, but a large part of the variation remained unshared, suggesting that classroom-based Mathematics grades only partly reflect the application and reasoning competencies assessed by the TKA.

These conclusions have implications for policy and practice. For SNBP, report scores and TKA results should be used as complementary sources of evidence, supported by procedures that monitor and calibrate differences in grading standards between schools. For schools and mathematics teachers, classroom assessment should include contextual and non-routine tasks that develop and capture conceptual understanding, application, problem solving, and reasoning, with grading criteria that focus on achievement. The study contributes early empirical evidence on assessment divergence under Indonesia's dual-assessment admission policy and introduces a simple quartile-based procedure that schools and policymakers can use to identify students whose report scores diverge from their standardized performance.

Future research is recommended to involve larger and more diverse samples, including applicants who were not admitted, to apply multilevel analysis that distinguishes within-school from between-school relationships, and to follow students into university to determine which combination of report scores and TKA results best predicts academic success. Studies of other TKA subjects and qualitative studies of teachers' grading practices would further clarify why the two forms of assessment diverge and how classroom assessment can be improved.

ACKNOWLEDGMENTS

The authors sincerely thank all those who supported the completion of this research. We also express our gratitude to the institutions that provided academic support, as well as to colleagues who provided constructive input to the successful implementation and completion of this research.

REFERENCES

Andina, E. (2025). Academic ability test (TKA) for quality oversight and accountability of national education. *Commission X: Education, Sports, Science, and Technology*, *17*, 15. Center for Parliamentary Analysis, Expertise Agency of the DPR RI.

Ani, Y. (2014). Penilaian autentik dalam Kurikulum 2013. In *Seminar Nasional Implementasi Kurikulum 2013* (pp. 742–749).

Arifin, Z. (2016). *Evaluasi pembelajaran*. PT Remaja Rosdakarya.

Arrafii, M. A. (2020). Grades and grade inflation: Exploring teachers' grading practices in Indonesian EFL secondary school classrooms. *Pedagogy, Culture & Society*, *28*(3), 477–499. https://doi.org/10.1080/14681366.2019.1663246

Asrijanty. (2014). Validitas prediktif bakat skolastik dan prestasi belajar sebagai kriteria seleksi masuk perguruan tinggi. *Jurnal Pendidikan dan Kebudayaan*, *20*(4), 515–534. https://doi.org/10.24832/jpnk.v20i4.163

Bejar, I. I., & Blew, E. O. (1981). Grade inflation and the validity of the Scholastic Aptitude Test. *American Educational Research Journal*, *18*(2), 143–156. https://doi.org/10.3102/00028312018002143

Berkhout, E., Pradhan, M., Rahmawati, Suryadarma, D., & Swarnata, A. (2024). Using technology to prevent fraud in high stakes national school examinations: Evidence from Indonesia. *Journal of Development Economics*, *170*, Article 103307. https://doi.org/10.1016/j.jdeveco.2024.103307

BSKAP Kemendikdasmen. (2025a). *Panduan pembelajaran dan asesmen pendidikan anak usia dini, jenjang pendidikan dasar, dan jenjang pendidikan menengah* (Edisi revisi ke-3). Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan Dasar dan Menengah. https://kurikulum.kemendikdasmen.go.id

BSKAP Kemendikdasmen. (2025b). *Tes Kemampuan Akademik (TKA)*. Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan Dasar dan Menengah. https://pusmendik.kemendikdasmen.go.id/tka/page/download

Chen, X., & Sanchez, E. I. (2024). *Has the predictive validity of high school GPA and ACT scores on postsecondary enrollment changed over time?* (ACT Research Report R2321). ACT. https://act.org/content/dam/act/unsecured/documents/pdfs/R2321-Has-the-Predictive-Validity-of-HSGPA-and-ACT-Scores-on-Postsecondary-Enrollment-Changed-2024-01.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Contreras, D. (2024). Gender differences in grading: Teacher bias or student behaviour? *Education Economics*, *32*(6), 762–785. https://doi.org/10.1080/09645292.2023.2252620

Ernawati, E., Habibah, R. Y., Syarifah, N., Firmansyah, F., & Attamimi, H. R. (2024). Item analysis test of science, Indonesian language, and mathematics using the Rasch model in elementary schools. *Jurnal Penelitian dan Evaluasi Pendidikan*, *28*(2), 195–209. https://doi.org/10.21831/pep.v28i2.75448

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Farozin, M., Astuti, B., Sanyata, S., Titi, D., Nurmalasari, E., & Sutanti, N. (2022). Rasch model for the need assessment instrument of academic guidance and counseling program in junior high school. *Jurnal Penelitian dan Evaluasi Pendidikan*, *26*(2), 177–185. https://doi.org/10.21831/pep.v26i2.51816

Ferdian, F. R., & Azwar, S. (2019). PAPS predictive validity in predicting the learning success of master of professional psychology students. *Jurnal Penelitian dan Evaluasi Pendidikan*, *23*(2), 117–128. https://doi.org/10.21831/pep.v23i2.25299

Ferman, B., & Fontes, L. F. (2022). Assessing knowledge or classroom behavior? Evidence of teachers' grading bias. *Journal of Public Economics*, *216*, Article 104773. https://doi.org/10.1016/j.jpubeco.2022.104773

Gunur, B., Makur, A. P., & Ramda, A. H. (2018). Hubungan antara kemampuan numerik dengan kemampuan pemecahan masalah matematis siswa di pedesaan. *MaPan: Jurnal Matematika dan Pembelajaran*, *6*(2), 148–160. https://doi.org/10.24252/mapan.2018v6n2a2

Hanna, R. N., & Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, *4*(4), 146–168. https://doi.org/10.1257/pol.4.4.146

Harun, H., Prayitno, P., Sudaryanti, S., Rolina, N., & Manaf, A. (2023). Karakteristik butir instrumen asesmen dinamis capaian karakter anak usia dini: Analisis Rasch model. *Jurnal Penelitian Ilmu Pendidikan*, *16*(2), 154–163. https://doi.org/10.21831/jpipfip.v16i2.63358

Hayat, B., Putra, M. D. K., & Suryadi, B. (2020). Comparing item parameter estimates and fit statistics of the Rasch model from three different traditions. *Jurnal Penelitian dan Evaluasi Pendidikan*, *24*(1), 39–50. https://doi.org/10.21831/pep.v24i1.29871

Irawan, W. (2026). *Gov't prioritizes inclusive access in 2026 student admissions system*. ANTARA News. https://en.antaranews.com/news/411516/govt-prioritizes-inclusive-access-in-2026-student-admissions-system

Jaehn, M., Hissbach, J., Frickhoeffer, M., Weppert, D., Zimmerhofer, A., Hampe, W., Kadmon, M., & Becker, N. (2025). Predictive validity of admission tests and educational attainment on preclinical academic performance: A multisite study. *BMC Medical Education*, *25*, Article 1255. https://doi.org/10.1186/s12909-025-07974-2

Juriyah, J., Fahrudin, T., & Jalaludin, J. (2025). Analisis kesiapan implementasi kebijakan Tes Kemampuan Akademik (TKA) 2025 oleh Kemendikbud. *YASIN: Jurnal Pendidikan dan Sosial Budaya*, *5*(6), 5708–5721. https://doi.org/10.58578/yasin.v5i6.7771

Kamelia, R. (2025). *Apakah hasil TKA memengaruhi nilai rapor siswa?* Tirto.id. https://tirto.id/apakah-hasil-tka-memengaruhi-nilai-rapor-siswa-cek-infonya-hh7a

Kristianti, K., & Handayani, I. (2023). Analisis literasi numerasi matematis peserta didik kelas XI dalam menyelesaikan soal asesmen kompetensi minimum. *JPMI (Jurnal Pembelajaran Matematika Inovatif)*, *6*(4), 1379–1390. https://doi.org/10.22460/jpmi.v6i4.17919

Lukmanulhakim, L. (2018). Nilai rapor sekolah menengah dan hubungannya dengan indeks prestasi komulatif mahasiswa. *Jurnal Visi Ilmu Pendidikan*, *10*(1), 1–7. https://doi.org/10.26418/jvip.v9i1.25958

Meeter, M. (2023). Predicting retention in higher education from high-stakes exams or school GPA. *Educational Assessment*, *28*(1), 1–10. https://doi.org/10.1080/10627197.2022.2130748

Ocy, D. R., Rahayu, W., & Makmuri, M. (2023). Development of a HOTS-based mathematical abstraction ability instrument in trigonometry using Riau Islands Province culture. *Jurnal Gantang*, *8*(1), 37–52. https://doi.org/10.31629/jg.v8i1.5654

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, *42*(3), 375–392. https://doi.org/10.3102/0162373720932188

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, *85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Saniah, S. L., & Nindiasari, H. (2023). Efektivitas flipped classroom diintegrasikan dengan model discovery learning terhadap kemampuan numerasi ditinjau dari disposisi matematis siswa SMA. *JPMI (Jurnal Pembelajaran Matematika Inovatif)*, *6*(1), 151–158. https://doi.org/10.22460/jpmi.v6i1.14472

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Setianingsih, W. L., Ekayanti, A., & Jumadi, J. (2022). Analisis kemampuan numerasi siswa SMP dalam menyelesaikan soal tipe Asesmen Kompetensi Minimum (AKM). *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(4), 3262–3273. https://doi.org/10.24127/ajpm.v11i4.5915

Sudjana, N. (2017). *Penilaian hasil proses belajar mengajar*. PT Remaja Rosdakarya.

Sulistiawan, C. H. (2016). Kualitas soal ujian sekolah matematika program IPA dan kontribusinya terhadap hasil ujian nasional. *Jurnal Penelitian dan Evaluasi Pendidikan*, *20*(1), 1–10. https://doi.org/10.21831/pep.v20i1.7516

Wibowo, S. B. (2025). *Menimbang ulang validitas nilai rapor dan tes akademik dalam PPDB Lampung 2025*. Universitas Muhammadiyah Metro. https://ummetro.ac.id/public/index.php/menimbang-ulang-validitas-nilai-raport-dan-tes-akademik-dalam-ppdb-lampung-2025

Yusnadi, Y., Lubis, Z., & Kamtini, K. (2021). The ability of national selection students to enter state universities (SNMPTN) in following the lecture process at the Faculty of Education, Universitas Negeri Medan. In *Proceedings of the 3rd International Conference on Innovation in Education, Science and Culture (ICIESC 2021)*. EAI. https://doi.org/10.4108/eai.31-8-2021.2313807

Zana, F. M., Sa'dijah, C., Susiswo, Anwar, L., & Zulnaidi, H. (2024). Curriculum and teacher assessment practices in mathematics learning: Alignment with higher order thinking skills in Indonesian secondary schools. *Journal on Mathematics Education*, *15*(4), 1311–1334. https://doi.org/10.22342/jme.v15i4.pp1311-1334

Zanga, G., & De Gioannis, E. (2023). Discrimination in grading: A scoping review of studies on teachers' discrimination in school. *Studies in Educational Evaluation*, *78*, Article 101284. https://doi.org/10.1016/j.stueduc.2023.101284

Zulaiha, S., Mulyono, H., & Ambarsari, L. (2020). An investigation into EFL teachers' assessment literacy: Indonesian teachers' perceptions and classroom practice. *European Journal of Contemporary Education*, *9*(1), 189–201. https://doi.org/10.13187/ejced.2020.1.189