Institutional Quality of Human Capital Assessment and Meritocratic Selection in a Post-Transition Higher Education Environment

Peter SZIKORA
https://orcid.org/0000-0001-8680-3880
Department of Business Sciences and Digital Skills
Keleti Károly Faculty of Business and Management
Obuda University, Budapest, Hungary
szikora.peter@kgk.uni-obuda.hu

Abstract

Reliable human capital assessment is an important institutional precondition for meritocratic selection and, indirectly, for productivity-enhancing resource allocation in post-transition economies. This study examines the institutional quality of student thesis evaluation in a post-transition higher education environment, with particular attention to whether grading patterns are associated with institutional proximity or hierarchical relationships between reviewers and the circle of academic advisors. The analysis is based on an anonymized administrative database from one university and covers 4,953 individual thesis reviews. Empirically, the paper combines review-level group comparisons, analyses of deviations from thesis-level average reviewer scores, and direct comparisons within a two-reviewer subsample. The results indicate a small but statistically detectable positive scoring shift associated with institutional proximity, while the effects of hierarchical relationships are not robust across specifications. At the same time, the overall magnitude of the observed relational effects remains limited, and no evidence of strong structural distortion is found. The findings suggest that the examined evaluation system is fundamentally

functional and relatively stable, although minor proximity-related effects cannot be fully excluded. By focusing on the micro-level institutional mechanisms through which performance is assessed and differentiated, the study contributes to the literature on human capital, institutional quality, and the meritocratic foundations of productivity-relevant selection in post-transition settings.

Keywords: *human capital assessment; institutional quality; evaluation system; institutional proximity; post-transition higher education environment*

Introduction

A key institutional precondition of productivity-led growth and long-term economic convergence is the reliable identification and assessment of human capital. In higher education, the evaluation of student performance is not merely an administrative or pedagogical task; it is also a mechanism through which skills, competencies, and professional potential are recognized, differentiated, and signaled. Therefore, the quality of assessment systems is linked to meritocratic selection, performance-based advancement, and the effective allocation of human resources.

This issue is particularly important in post-transition settings, where long-term economic transformation depends not only on market liberalization and macroeconomic stabilization, but also on the credibility and consistency of formal institutions. Institutions involved in the formation and assessment of human capital are especially relevant, because they shape how reliably qualifications and competencies become visible and comparable. If evaluation systems are affected by relational factors or institutional bias, this may weaken both organizational trust and the informational value of human capital assessment.

From this perspective, higher education evaluation systems can be understood as empirically observable manifestations of institutional quality. A particularly relevant question is whether grading remains independent from organizational and hierarchical relations among academic actors. Institutional proximity, shared organizational embeddedness, or rank differences may influence judgment even in the absence of explicit favoritism. Assessing whether such effects are present is therefore important for understanding how consistently human capital is evaluated in an institutional environment expected to support performance-based differentiation.

This study examines these issues through the grading of student theses in a post-transition higher education environment. More specifically, it investigates whether reviewers' scores are associated with their institutional proximity to the circle of academic advisors and with the hierarchical relationship between

reviewers and advisors. The empirical analysis is based on an anonymized administrative database from one university and covers 4,953 individual thesis reviews. To identify potential relational effects, the study combines review-level group comparisons, analyses of deviations from thesis-level average reviewer scores, and direct comparisons within a two-reviewer subsample.

The results show a small but statistically detectable positive shift associated with institutional proximity, whereas the role of hierarchical relationships is not robust across specifications. At the same time, the magnitude of the observed effects remains limited, and the overall evaluation system does not display evidence of strong structural distortion. These findings suggest that the examined review system is fundamentally functional and relatively stable, even though minor proximity-related effects cannot be fully excluded.

The contribution of the study is twofold. First, it provides empirical evidence on the institutional functioning of human capital assessment in higher education in a post-transition context. Second, it shows that the reliability of performance assessment may be partly conditioned by organizational relations, making the analysis relevant not only for educational governance, but also for broader discussions of meritocratic selection, human capital signaling, and productivity-relevant allocation.

Literature Review

The literature on productivity-driven growth and economic convergence widely emphasizes that human capital and institutional quality interact closely in shaping development trajectories. Rodríguez-Pose and Ganau (2022) argue that productivity differences cannot be understood solely through variations in resources, investment, or factor endowments, because the institutional environment itself influences the efficiency of economic performance. Sanga, Ahouakan, and Adje (2022) reach a similar conclusion in their analysis of labor productivity and institutional quality, showing that the stability and reliability of formal institutional frameworks are directly linked to economic performance. In the same broader direction, Ngqoleka et al. (2025) emphasize that human capital, technological investment, and institutional quality jointly contribute to greater economic complexity and development potential. Taken together, these approaches suggest that human capital should not be treated as an isolated stock of capabilities; rather, its economic relevance is closely connected to the institutional mechanisms that shape its formation, evaluation, and selection.

One important implication of these studies is that the economic significance of human capital depends not only on its quantitative presence, but also on how skills, knowledge, and performance become recognizable, comparable, and reliably

assessable within institutional settings. If the reliability of formal institutions contributes to productivity and development, then it becomes equally important to examine the specific mechanisms through which human capital is assessed and certified. The question is not merely whether individuals possess competencies, but whether institutional systems are capable of identifying and differentiating them in a consistent and performance-based manner.

The relationship between human capital and productivity has been confirmed by numerous studies at both the firm and regional levels. Backman (2014) found that human capital accumulation is positively associated with productivity at both levels, suggesting that the contribution of skills and knowledge extends to broader economic performance. Marrocu, Paci, and Usai (2022) similarly argue that universities contribute to regional productivity both directly and indirectly. If higher education is a key institutional setting in which human capital is formed and certified, then the reliability of the internal assessment mechanisms operating within it cannot be regarded as a purely administrative issue. Rather, it is linked to the credibility of human capital identification itself.

In this sense, higher education performs a dual function. On the one hand, it contributes to the development of knowledge, skills, and professional competencies. On the other hand, it evaluates and differentiates these in institutionalized forms. The evaluation of student performance is therefore not merely an internal pedagogical practice, but also a mechanism that generates information about performance and indirectly shapes selection, progression, and later career trajectories. If this assessment process is affected by bias or context-sensitive influences, it may weaken the informational value of the performance signals that higher education institutions produce.

This broader connection is also reflected in more direct literature on the economic role of higher education. Nikolli and Shehu (2022), in analyzing the links between education and economic growth, argue that the performance of education systems is decisive in the long run for economic output and development capacities. Among studies published in Economicus, Osmënaj (2023) examined the conditions for the success of projects at higher education institutions, thereby incorporating institutional performance and efficiency considerations into the analysis. Qosja et al. (2020) emphasized the role of perceived institutional quality in evaluating the entrepreneurial environment, a point conceptually linked to the broader question of whether formal institutions are capable of ensuring fair, predictable, and credible operation. Pere and Minxhozi (2011), in turn, examined the funding structure of higher education and showed that the operating conditions of higher education institutions can be understood within a broader institutional framework. Although these studies do not specifically focus on student thesis evaluation, they reinforce the view that the internal mechanisms of higher education institutions, including their

assessment and grading systems, are relevant to the broader institutional quality of human capital development.

The macro-level literature on human capital and institutional quality therefore supports the claim that assessment mechanisms are economically relevant. At the same time, this literature typically remains at an aggregate level. It links the performance of countries, regions, firms, or education systems to institutional factors, but offers less insight into how the specific, micro-level mechanisms of human capital evaluation actually function in practice. Particularly underexplored is the question of to what extent decisions within higher education assessment can be regarded as strictly performance-based, and to what extent they are shaped by organizational and hierarchical relations among the actors involved.

The literature on peer review and evaluation bias provides a relevant analytical framework for this purpose. Lee, Sugimoto, Zhang, and Cronin (2012), in their influential review, show that peer review processes may be susceptible to bias at several points, especially when reviewers are exposed to information related to the author, affiliation, or institutional background. Fox, Meyer, and Aimé (2023) also demonstrate empirically that the structure of the review process, including the use of double-blind reviewing, can affect reviewer evaluations and editorial decisions. These findings suggest that reviews do not merely reflect the quality of the work under evaluation, but may also be shaped by institutional and procedural conditions.

Affiliation-related effects are particularly relevant for the present study. Von Wedel et al. (2024) specifically examined affiliation bias and found that organizational background can influence evaluation outcomes. Although their study focuses on the evaluation of scholarly work rather than student theses, the underlying structural problem is closely related to the one examined here: whether the evaluator's judgment primarily reflects the quality of the work, or whether organizational proximity also enters into the decision. From this perspective, the peer review literature is valuable not as a direct empirical equivalent, but as a structurally relevant body of scholarship that helps explain why institutional proximity may matter in the evaluation of student theses as well.

At the same time, an important distinction must be maintained between the two contexts. Journal peer review is primarily concerned with the publishability of academic manuscripts, whereas higher education evaluation is an internal assessment mechanism more directly linked to the formal appraisal of student performance. The two fields therefore cannot be treated as identical. Nevertheless, the structural logic of possible biases is comparable. In both contexts, formal expert judgment may intersect with the effects of affiliation, institutional embeddedness, or organizational proximity.

The literature on higher education evaluation systems further supports this line of inquiry. Fischer and Hänze (2018) examined the extent to which higher

education evaluations reflect actual performance and the extent to which they may be distorted by other factors. Goos and Salomons (2017), in the context of course evaluations, showed that selection biases can significantly alter the interpretation of observed evaluation outcomes. Sætre (2024), meanwhile, argues that fairness is not merely a normative expectation, but a socially emerging and continually renegotiated practice within evaluation processes. These studies collectively suggest that the analysis of assessment systems cannot be limited to final scores alone; it must also take into account the institutional and relational conditions under which those scores are produced.

This insight is highly relevant for the present study. If fairness is not simply an abstract principle but an institutional practice, then the evaluation of student theses cannot be reduced to a binary distinction between "correct" and "incorrect" decisions. A more appropriate question is whether the grading system is capable of consistently upholding the norm of performance-based evaluation in an environment where actors are organizationally connected and status relationships may matter. From this perspective, even relatively small relational effects are analytically important, because they illuminate finer dimensions of institutional quality that would otherwise remain hidden.

Taken together, the above literature provides a strong starting point while also revealing a clearly defined research gap. Studies on human capital, productivity, and institutional quality emphasize the importance of reliable institutions, but generally do not directly examine the micro-level evaluation processes through which human capital is assessed. The literature on peer review and evaluation fairness investigates bias and distortion in considerable detail, but usually does not connect these issues to the economic significance of human capital quality or to the specific institutional setting of post-transition higher education. Studies on the economic role of higher education underline the importance of universities for productivity and development, yet they pay comparatively little attention to whether concrete assessment mechanisms of student performance can be considered robust, meritocratic, and relatively neutral.

The research gap is therefore simultaneously one of scale, context, and institutional focus. The macro-level literature on institutional quality rarely moves down to specific evaluation practices, while the peer review literature typically examines distortion within the publication process rather than within the institutional setting of human capital assessment in higher education. This study addresses that gap by examining the evaluation of student theses as one of the institutional mechanisms through which human capital is evaluated.

The central research question is whether reviewers' scores are associated with institutional proximity to the circle of advisors, and whether they are related to the hierarchical relationship between the reviewer and the advisors. If evaluations primarily reflect the quality of the theses, then the evaluation system

supports meritocratic selection. If, however, evaluation is systematically linked to organizational or hierarchical proximity, then the institutional reliability of human capital assessment becomes at least partly questionable. The significance of this question extends beyond the technical functioning of an evaluation procedure; it concerns the extent to which higher education institutions are able to generate and communicate credible performance signals.

At the same time, the existing literature does not necessarily lead one to expect strong or dramatic distortions. The findings of Lee et al. (2012), Fox et al. (2023), and von Wedel et al. (2024) rather suggest that such effects are often limited in magnitude, yet still empirically detectable. Accordingly, the present study does not seek to provide a sweeping normative judgment about the review system as a whole. Instead, it examines whether the system exhibits mild but systematic deviations that can be associated with institutional proximity or hierarchical embeddedness.

This approach lies at the intersection of several strands of literature. The economic literature on human capital and institutional quality (Backman, 2014; Marrocu et al., 2022; Rodríguez-Pose & Ganau, 2022) provides the theoretical basis for understanding why performance assessment can be considered an institutional issue relevant to both productivity and selection. Studies on the operation of higher education institutions (Nikolli & Shehu, 2022; Osmënaj, 2023; Pere & Minxhozi, 2011; Qosja et al., 2020) contribute to an understanding of higher education as an institutional space structured by organizational decisions and evaluation mechanisms. Research on evaluation fairness and biases in peer review and assessment (Fischer & Hänze, 2018; Fox et al., 2023; Goos & Salomons, 2017; Lee et al., 2012; Sætre, 2024; von Wedel et al., 2024) shows that assessment outcomes are shaped not only by the performance under evaluation, but also by the relational and institutional characteristics of the evaluation setting. By bringing these strands together, the present study interprets the evaluation of student theses as a specific and empirically testable institutional mechanism of human capital assessment in a post-transition higher education environment.

Research Questions and Hypotheses

The empirical objective of the study is to assess whether the grading of student theses exhibits systematic relational effects associated with institutional proximity or hierarchical embeddedness. Consistent with the literature reviewed above, the analysis treats higher education assessment as a micro-level institutional mechanism of human capital evaluation. The focus is therefore not on bias in a general sense, but on whether mild yet systematic deviations in scoring can be linked to organizational and status-related relationships.

The first research question examines whether institutional proximity is associated with a shift in reviewers' scoring.

Q1. Is institutional proximity between the reviewer and the advisory group associated with systematic positive or negative deviation from the thesis-level average reviewer score?
H1. Reviewers who share institutional affiliation with at least one advisor associated with a given thesis tend to assign slightly higher scores relative to the thesis-level average reviewer score than reviewers who do not have such institutional proximity.

The second research question examines whether hierarchical relationships influence the direction of scoring.

Q2. Is the reviewer's deviation from the thesis-level average reviewer score associated with hierarchical position relative to the highest-ranking advisor?
H2. Hierarchical position may be associated with directional differences in reviewer scoring, with lower-ranking reviewers expected to evaluate somewhat more favorably relative to the thesis-level average.

The third research question concerns the magnitude of reviewer discrepancy.

Q3. Do institutional proximity and hierarchical relations affect the magnitude, rather than only the direction, of reviewer deviation?
H3. Institutional proximity and ties to the advisor's organizational background do not produce a substantial difference in the magnitude of reviewer variation; rather, these factors are more likely to slightly affect the direction of scoring without undermining the overall consistency of the evaluation system.

The fourth research question focuses on two-reviewer theses.

Q4. In two-reviewer theses, does the reviewer institutionally closer to the advisory group assign a higher score than the reviewer without such proximity?
H4. In cases where only one of the two reviewers is affiliated with the same institution as at least one advisor, the reviewer with that institutional connection assigns a higher average score than the other reviewer.

Overall, the hypotheses do not assume that the grading system is strongly biased. Rather, they examine whether organizational and hierarchical proximity exert mild yet empirically detectable effects on evaluation. In this way, the study assesses whether thesis evaluation operates as a relatively stable and meritocratic mechanism of human capital assessment.

Methodology

The study is based on a fully anonymized university-level administrative database containing evaluations of student theses. The analysis uses a review-level dataset, meaning that each row represents one individual review linked to a specific thesis. The final analytical sample consisted of 4,953 reviews. The database provided linked information on reviewers' institutional affiliations, the institutional affiliations of the advisors associated with each thesis, and the review scores assigned. Only anonymized identifiers were used during data processing, and the analysis did not involve any personally identifiable information.

The review-level approach follows directly from the research question. The purpose of the study is not to explain the average grades of theses, but to examine how individual reviewers score relative to other reviewers of the same thesis and whether such differences are associated with institutional or hierarchical relationships. Accordingly, the main unit of analysis is the review as an act of evaluation rather than the thesis as a completed product.

Only records meeting the criteria of data quality and comparability were retained. Archived evaluations, non-final evaluations, and cases without clearly identifiable reviewers were excluded. When establishing advisory background, only active advisory relationships were considered.

An examination of the review structure showed that the overwhelming majority of theses involved two reviewers. In the review-level dataset, 4,440 reviews belonged to two-reviewer theses, accounting for 89.6% of all observations. By comparison, single-reviewer cases represented 2.9%, three-reviewer cases 7.2%, and four-reviewer cases only 0.2%. Because the two-reviewer structure clearly dominated the database, the empirical analysis combines review-level models estimated on the full sample with a separate subsample analysis restricted to two-reviewer theses. This subsample contained 2,220 theses and made direct within-thesis reviewer comparisons possible.

The empirical strategy is based on the premise that relational effects are more meaningfully detected through relative differences within theses than through raw scores alone. In addition to the original review score, we therefore calculated the average score for each thesis and derived a signed deviation variable indicating how far a given reviewer's score deviated from the thesis-

level average. This indicator is directly linked to the first and second research questions, because it allows us to test whether institutional proximity or hierarchical relationships are associated with more positive or more negative scoring relative to the same thesis.

Using deviation from the thesis-level average is methodologically justified because theses differ in quality. A comparison based only on raw scores would conflate thesis quality with reviewers' scoring tendencies. By contrast, deviation from the thesis-level average captures whether a reviewer assigned a more lenient or stricter score relative to the other reviewer(s) of the same thesis.

In addition, we calculated the absolute deviation from the thesis-level average, which captures the magnitude of reviewer deviation irrespective of direction. While signed deviation indicates whether reviewers score above or below the thesis-level average, absolute deviation shows how far they deviate from that average. This distinction makes it possible to assess separately whether relational factors influence the direction of scoring or the overall consistency of reviewer behavior.

In this study, the term "institution" is used in a formal-organizational sense. It refers to the officially recorded organizational unit or affiliation within the university structure to which reviewers and advisors belong. Accordingly, institutional proximity was operationalized by identifying, for each review, whether at least one active advisor associated with the thesis had the same institutional affiliation as the reviewer. This was coded as a binary variable indicating the presence or absence of direct institutional proximity between the reviewer and the advisory group. We also considered the number of advisors from the same institution, allowing institutional proximity to be interpreted not only as a dichotomous condition but also in terms of intensity. Although this measure cannot capture all dimensions of informal relationships, personal acquaintance, or professional networking, formal institutional affiliation provides an appropriate starting point for examining whether organizational proximity is associated with systematic differences in scoring.

To operationalize hierarchical relationships, we constructed a rank variable based on the positions of reviewers and advisors. In theses with multiple advisors, the reviewer's rank was compared to that of the highest-ranking advisor associated with the thesis. On this basis, a categorical variable was created indicating whether the reviewer occupied a lower, equal, or higher rank relative to the highest-ranking advisor. This procedure allowed the second research question to be tested empirically by examining whether status differences are associated with systematic scoring tendencies.

The comparison to the highest-ranking advisor is a simplifying assumption, since actual organizational relations may be more complex than can be fully represented by an ordered rank structure. Nevertheless, it provides an analytically

consistent way to make the direction and basic structure of formal status differences empirically observable.

To test the hypotheses, the analysis proceeded in several stages. First, descriptive statistics were used to map the structure of the review process and the main scoring patterns. Second, two-sample t-tests and analysis of variance were applied to examine whether mean differences in scoring were associated with institutional proximity or hierarchical position. Third, linear regression models were estimated to assess the effects of relational variables within a multivariate framework. This cumulative analytical strategy makes it possible to examine the same problem from descriptive, comparative, and multivariate perspectives.

In addition to the review-level analyses, a separate direct comparison was conducted on the two-reviewer subsample. More specifically, we examined those theses in which exactly one of the two reviewers shared institutional affiliation with the advisory circle. This subset contained 702 theses and allowed a direct test of whether the reviewer with institutional proximity assigned a systematically higher score to the same thesis.

The advantage of this within-thesis comparison is that it controls for several important unobserved sources of heterogeneity. Within the same thesis, both reviewers evaluate the same underlying performance, so observed score differences are less likely to be attributable to variation in thesis quality. Although this design does not establish causality in a strict sense, it provides a stronger internal comparison than a simple between-group test estimated on the pooled sample.

The methodological aim is therefore to determine whether the grading system exhibits mild yet systematic deviations relevant to the institutional reliability of human capital assessment.

Results

In the first step of the analysis, we examined the basic structure of the review dataset. The review-level analytical sample contained 4,953 individual reviews. The vast majority of the reviews were associated with theses evaluated by two reviewers, which confirmed that deviation from the thesis-level average reviewer score is an appropriate dependent variable for comparing reviewer behavior. As expected, the mean signed deviation in the full sample was 0, while the standard deviation was 4.66 points, indicating substantial differences between the reviews of the same thesis.

TABLE 1. Distribution of the number of reviewers in the review-level sample

Number of reviewers per thesis	Frequency	Percentage (%)
1	144	2.9
2	4440	89.6
3	357	7.2
4	12	0.2
Total	4,953	100.0

Based on Table 1, the dominance of two-reviewer theses is clear: nearly nine-tenths of all observations fell into this category. This justified conducting a separate robustness analysis on the two-reviewer subsample alongside the analyses based on the full sample. Regarding other characteristics of the sample, 51.6% of reviewers had at least one advisor affiliated with the same institution as the reviewer, while 48.4% did not have such a connection. In addition, 20.1% of theses had multiple advisors. With respect to hierarchical relationships, 46.0% of reviewers were of lower rank than the highest-ranking member of the advisory group, 24.2% were of the same rank, and 29.8% were of higher rank.

In the next step, we examined whether institutional proximity was associated with deviation from the thesis-level average reviewer score. Group means calculated for the full sample showed that reviewers who had at least one advisor from the same institution deviated slightly in a positive direction from the thesis-level average, whereas reviewers without such an institutional connection showed a slight negative deviation.

TABLE 2. Institutional proximity and deviation from the thesis-level average reviewer score

Institutional connection	N	Average deviation from the thesis-level average reviewer score	Standard deviation
No advisor from the same institution	2399	-0.138	4.673
At least one advisor from the same institution	2554	0.130	4.644
Total	4953	0.000	4.659

The difference shown in Table 2 was statistically significant based on a two-sample t-test ($t = -2.021$; $p = 0.043$), but the effect size remained very small (Cohen's $d \approx -0.058$). This indicates that the direction of the association related to institutional proximity is detectable, but its substantive magnitude is limited. At the level of the evaluation system as a whole, therefore, the results do not suggest a strong bias, but rather a slight positive shift associated with institutional proximity. By contrast, the analysis of absolute deviation did not reveal a significant difference between the two groups ($p = 0.523$), indicating that institutional proximity was

not associated with reviewers deviating more or less strongly than others. This suggests that the relationship may affect the direction of scoring rather than the magnitude of reviewer deviation.

A similar pattern emerged with respect to the hierarchical relationship between the reviewer and the advisory group, although its strength remained limited. Lower-ranking reviewers showed a slightly positive deviation on average, whereas higher-ranking reviewers showed a slightly negative deviation relative to the thesis-level average.

TABLE 3. Hierarchical relationship and deviation from the thesis-level average reviewer score

Rank	N	Average deviation	Standard deviation
Reviewer of lower rank	2279	0.137	4.672
Same rank	1200	-0.017	4.502
Reviewer of higher rank	1474	-0.198	4.760
Total	4953	0.000	4.659

The pattern shown in Table 3 is consistent with the hypothesis that lower-status reviewers evaluate somewhat more cautiously, whereas higher-status reviewers evaluate more autonomously. However, the analysis of variance did not reveal a statistically significant difference between the groups ($F = 2.330$; $p = 0.097$). The effect of rank therefore appeared as a tendency rather than a stable and robust association. The analysis of absolute deviation also yielded no significant result ($p = 0.908$), meaning that hierarchical relationship did not substantially affect the magnitude of reviewer deviation. Overall, this suggests that status differences may be present at most as a weak directional effect, but do not undermine the overall consistency of the system. In the subsequent analysis, we estimated linear regression models to examine how institutional proximity, hierarchical relationships, and multi-advisor structure were jointly associated with reviewer deviation. The model explaining signed deviation from the thesis-level average reviewer score was statistically significant overall ($F = 2.931$; $p = 0.032$), although the explained variance remained very low ($R^2 = 0.002$).

TABLE 4. Linear regression model for deviation from the thesis-level average reviewer score

Explanatory variable	B	Standard error	Beta	p
At least one advisor from the same institution	0.266	0.133	0.029	0.045
Rank	-0.169	0.078	-0.031	0.031
Thesis with multiple advisors	-0.080	0.167	-0.007	0.634

According to Table 4, institutional proximity continued to show a significant but modest positive association with reviewer deviation. This confirms the pattern observed in the previous group comparisons: reviews associated with institutional proximity to the advisory group were, on average, slightly more favorable relative to the thesis-level average. The negative coefficient for rank suggests that higher-ranking reviewers tended to deviate downward from that average. At the same time, given the overall structure of the results, this effect cannot be regarded as particularly robust. The presence of multiple advisors did not prove to be a relevant factor in this model. Importantly, the very low R^2 value indicates that these relational variables explain only a negligible share of reviewer-level variation; therefore, the statistically significant coefficients should be interpreted as weak directional associations rather than substantively strong predictors. By contrast, the regression model estimated for absolute deviation was not significant in the full sample ($p = 0.212$), which further reinforces the conclusion that institutional and hierarchical factors affect primarily the direction, rather than the magnitude, of reviewer deviation.

A separate analysis of the two-reviewer subsample largely confirmed the pattern observed in the full sample. In this subsample, the average deviation associated with institutional proximity persisted: reviewers who had at least one advisor from the same institution assigned, on average, higher scores relative to the thesis-level average than reviewers without such a connection. The difference was statistically significant in this subsample as well ($t = -2.048$; $p = 0.041$), whereas the effect of rank remained non-significant. This suggests that the institutional proximity effect observed in the full sample is not merely a by-product of more complex review structures, but is also visible within the most common thesis type.

The strongest direct test was provided by the subset of two-reviewer theses in which exactly one of the two reviewers shared institutional affiliation with the advisory group. This subset contained 702 theses. In this comparison, the mean directed difference between the score assigned by the reviewer with institutional proximity and that assigned by the other reviewer was 0.86 points, which differed significantly from zero ($t = 2.572$; $p = 0.010$). This indicates that, within the same thesis, the reviewer with institutional proximity assigned a slightly higher score on average. At the same time, this reviewer did not assign higher scores significantly more often than the other reviewer (binomial test: $p = 0.473$; comparison of positive and negative signs: $p = 0.145$). The effect therefore does not appear as strong favoritism consistently operating in the same direction in every case, but rather as a modest tendency toward somewhat higher average scores.

Overall, the results suggest that no strong structural bias is observable in the evaluation system under examination. A slight positive shift associated with institutional proximity was detectable across several analyses, but the magnitude of this effect remained consistently small. By contrast, the effects related to

hierarchical relationships did not prove sufficiently stable. More generally, the relational factors examined here did not provide a meaningful explanation for the magnitude of reviewer deviation. Taken together, these findings suggest that relational effects cannot be fully ruled out, but their strength is limited and does not indicate a deeper structural distortion in the functioning of the evaluation system. From the perspective of the institutional quality of human capital assessment, this supports a dual interpretation: organizational proximity is not entirely irrelevant to evaluation outcomes, yet the basic functionality and relative stability of the examined system appear to remain intact.

Discussion

Overall, the findings suggest that the thesis evaluation system under study is fundamentally stable and functional, although minor shifts in scoring may occur under certain relational conditions. The most consistent pattern concerns institutional proximity: when the reviewer and the advisory group share the same institutional affiliation, scores tend to shift slightly in a more favorable direction on average. This pattern appears across several analyses, including full-sample group comparisons, regression estimates, and direct comparisons within the two-reviewer subsample. At the same time, the magnitude of the effect remains consistently small. Institutional proximity therefore does not appear to override performance-based evaluation, but may generate a limited positive shift in scoring.

On this basis, it would not be justified to characterize the evaluation system as a whole as biased or structurally distorted. A more accurate interpretation is that evaluation mechanisms embedded in formal institutions are not fully independent of organizational context, yet in the present case these effects remain limited in magnitude. The results suggest that organizational proximity is not irrelevant, but neither does it emerge as a dominant organizing principle of the system.

From a broader institutional perspective, this is a meaningful finding. Empirical assessments of institutional quality do not necessarily depend on identifying dramatic failures; they may also concern smaller tensions between formal rules and actual practice. In this sense, the results suggest that the formally performance-based logic of the evaluation system coexists with subtle relational effects. This does not invalidate the meritocratic character of the system, but rather qualifies it: meritocratic evaluation need not imply absolute relational neutrality if relational effects remain limited and non-dominant.

It is particularly relevant that institutional proximity was not systematically associated with the magnitude of reviewer deviation. This suggests that such relationships are not linked to general inconsistency, unpredictability, or more extreme evaluation behavior. Rather, the results indicate a small directional shift

in scoring in certain situations. If relational effects increased the dispersion or instability of scores, that would point to deeper institutional weakness. Here, by contrast, the findings suggest that the internal structure of the evaluation system remains broadly intact.

The results for hierarchical relationships were weaker and less robust. Although some analyses indicated a tendency for lower-ranking reviewers to deviate upward and higher-ranking reviewers to deviate downward relative to the thesis-level average, these patterns were not consistently significant. This suggests that the role of hierarchy is likely more complex than can be captured by the rank variables used here. Hierarchical effects may operate in more context-dependent ways, may interact with other organizational characteristics, or may be expressed more strongly in informal or communicative dimensions of evaluation than in scores themselves.

The findings should therefore be interpreted with caution. Several associations were statistically significant, but their effect sizes were small. Statistical detectability and practical significance are not identical in this case. The results point to subtle institutional sensitivities rather than strong bias mechanisms. The small size of the effect does not make the phenomenon irrelevant, but it does mean that the substantive impact of institutional proximity on evaluation outcomes remains modest.

From the perspective of human capital assessment, this can be seen as a relatively favorable outcome. The evaluation of student performance does not become a direct function of organizational relationships. In a post-transition institutional environment, this is particularly important, because meritocratic evaluation is a prerequisite for the credible identification of skills, competencies, and performance. If formal evaluation mechanisms are only moderately sensitive to relational proximity, the system remains capable of fulfilling a core institutional function in the differentiation and assessment of human capital.

The study also has several limitations. First, the analysis is based on a single institutional setting, which limits generalizability. It remains unclear to what extent similar patterns would appear in other universities, under different organizational cultures, or under alternative review arrangements. Second, the "true" quality of theses cannot be observed directly. The analysis therefore does not determine whether specific scores are closer to an external quality standard, but rather whether reviewers exhibit systematic relative differences in scoring. Third, institutional proximity and hierarchical relationships could only be operationalized through available administrative variables. These indicators capture certain formal relational structures, but they cannot fully measure informal ties, prior collaboration, personal acquaintance, or professional networks. Fourth, the review-level design implies that observations are not fully independent, because multiple reviews are nested within the same theses. Although the use of

thesis-level average deviations partly addresses between-thesis heterogeneity, the regression models do not explicitly model this nested structure. Future research could therefore apply multilevel models or thesis-clustered standard errors to estimate relational effects while accounting more directly for the dependence of observations within theses.

These limitations also point to directions for future research. Multi-institutional studies would help determine whether the patterns identified here are general or context-specific. It would also be useful to incorporate additional indicators of thesis quality beyond reviewer scores. In addition, qualitative or mixed-methods approaches could shed light on the informal norms, expectations, and organizational routines that shape evaluation behavior, as well as on how reviewers themselves interpret fairness, independence, and professional loyalty.

Overall, the findings suggest that organizational proximity is detectable in the evaluation system under study, but does not indicate serious or systemic distortion. The system can therefore be described as a fundamentally functional and relatively robust institutional mechanism in which performance-based evaluation remains dominant despite the presence of limited relational effects. From the perspective of the institutional quality of human capital assessment, this is a cautiously positive conclusion.

Conclusion

The aim of this study was to examine whether systematic discrepancies in the evaluation of student theses can be linked to institutional proximity or hierarchical relationships between reviewers and the advisory group. The results provide no evidence of a strong system-level bias in the evaluation system under examination. Although institutional proximity was associated with a slight positive shift in scoring across several analyses, the magnitude of this effect remained consistently small and did not indicate deeper structural distortion in the grading process. By contrast, hierarchical relationships did not prove sufficiently stable to be interpreted as a robust explanatory factor. The direct comparison of two-reviewer theses further confirmed that institutional proximity may confer a modest scoring advantage. At the same time, this effect was limited in size and did not appear as consistent favoritism across all cases. Taken together, the findings suggest that the evaluation system is fundamentally functional and relatively robust: even in the presence of organizational embeddedness, it appears to preserve the core features of merit-based assessment. In a broader sense, the study highlights the importance of the institutional quality of human capital assessment. In settings where the reliable identification of skills, competencies, and performance matters for productivity-oriented development and long-term convergence, a relatively

stable and only moderately context-sensitive evaluation system may be regarded as a favorable institutional characteristic. For this reason, the analysis of higher education evaluation mechanisms should be understood not only as a matter of educational organization, but also as a question of broader economic relevance related to human capital quality and the reliability of formal institutional processes.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this work, the author used generative AI-assisted tools in order to support language refinement, proofreading, and the improvement of textual clarity. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the final content of the publication.

References

Backman, M. (2014). Human capital in firms and regions: Impact on firm productivity. Papers in Regional Science, 93(3), 557–575. https://doi.org/10.1111/pirs.12005

Fischer, E., & Hänze, M. (2018). Bias hypotheses under scrutiny: Investigating the validity of student assessment of university teaching by means of external observer ratings. Assessment & Evaluation in Higher Education, 44(5), 772–786. https://doi.org/10.1080/02602938.2018.1535647

Fox, C. W., Meyer, J., & Aimé, E. (2023). Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. Functional Ecology, 37(5), 1144–1157. https://doi.org/10.1111/1365-2435.14259

Goos, M., & Salomons, A. (2017). Measuring teaching quality in higher education: Assessing selection bias in course evaluations. Research in Higher Education, 58(4), 341–364. https://doi.org/10.1007/s11162-016-9429-8

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2012). Bias in peer review. Journal of the Association for Information Science and Technology, 64(1), 2–17. https://doi.org/10.1002/asi.22784

Marrocu, E., Paci, R., & Usai, S. (2022). Direct and indirect effects of universities on European regional productivity. Papers in Regional Science, 101(5), 1105–1133. https://doi.org/10.1111/pirs.12698

Ngqoleka, S., Ncanywa, T., Mpongwana, Z., & Asaleye, A. J. (2025). Industrial diversification in emerging economies: The role of human capital, technological investment, and institutional quality in promoting economic complexity. Sustainability, 17(15), Article 7021. https://doi.org/10.3390/su17157021

Nikolli, E., & Shehu, E. (2022). Relationship between education and economic growth. *Economicus*, *21*(1), 100–115. https://doi.org/10.58944/uqli8167

Osmënaj, J. (2023). Projects' success and respective factors that produce it: The case of higher education institutions in Albania. Economicus, 22(1), 95–114. https://doi.org/10.58944/fjjy1989

Pere, E., & Minxhozi, L. (2011). Financing Albanian higher education: growth between the public and private sectors. European Journal of Higher Education, 1(2–3), 202–219. https://doi.org/10.1080/21568235.2011.596668

Qosja, E., Keci, I., & Pirani, E. (2020). Evaluating entrepreneurship framework through perceived institutional quality. *Economicus, 19*(2), 76–95. https://doi.org/10.58944/qxda6807

Rodríguez-Pose, A., & Ganau, R. (2022). Institutions and the productivity challenge for European regions. Journal of Economic Geography, 22(1), 1–25. https://doi.org/10.1093/jeg/lbab003

Sætre, C. (2024). Behind the grades: Co-constructing fairness to reach agreement in evaluative judgement. Assessment & Evaluation in Higher Education, 50(2), 236–249. https://doi.org/10.1080/02602938.2024.2373789

Sanga, D., Ahouakan, E. W. V., & Adje, N. A. F. (2022). Institutional quality and labor productivity: Evidence from the Economic Community of West African States countries. Journal of Economic Integration, 37(3), 540–557. https://doi.org/10.11130/jei.2022.37.3.540

von Wedel, D., Schmitt, R. A., Thiele, M., Leuner, R., Shay, D., Redaelli, S., & Schaefer, M. S. (2024). Affiliation bias in peer review of abstracts by a large language model. JAMA, 331(3), 252–253. https://doi.org/10.1001/jama.2023.24641