

Do local curriculum scores correlate with national residency test results? A pluriannual, nationwide survey of Italian Medical Universities

Livio Tarchi,¹ Stefano Damiani,¹ Simone Marini,² Claudio Cappelli,³ Giammaria Liuzzi,⁴ Massimo Minerva,⁵ Pierluigi Politi¹

¹Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; ²Department of Epidemiology, University of Florida, Gainesville, FL, USA; ³Primary Care, ASUR MARCHE - AV5, San Benedetto del Tronto (AP), Italy; ⁴Department of Clinical Chemistry, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy; ⁵Department of Anesthesiology and Intensive Care, ASST Melegnano e della Martesana Sede di Cernusco, Cernusco Sul Naviglio (MI), Italy

ABSTRACT

In Italy, to match medical doctors to specialty residencies, the Minister of Education organizes a national multiple choice questions exam composed of 140 items. The curriculum grants additional points based on the candidates' grade average and graduation mark. Both national and international policies assume that curriculum scores can be evaluated in absolute terms, irrespective of the course attended by the candidate or its University of graduation. A fair inquiry over the current criteria of selection is not available to this day. Hence, this study aimed at investigating the degree of association between results in the Italian national residency test and the score given for each candidate's curriculum. We calculated Spearman's correlation coefficients between the score obtained at the 2019 test and curriculum scores. A moderate/high degree of correlation ($\rho=0.600$) was reached. Spearman's ρ ranged from 0.447 to 0.788, considering the single Universities. We thus measured a modest but non-negligible between-University variability. As differences between test and curriculum scores across institutions may determine the admission or not to the residency for many candidates, the authors call for broader attention on the topic.

Correspondence: Livio Tarchi, Department of Brain and Behavioral Sciences, University of Pavia, via Bassi 21, 27100 Pavia, Italy.

Tel.: +39.0382.986456.

E-mail: livio.tarchi01@universitadipavia.it

Key words: Testing; curriculum evaluation; postgraduate training; admission; selection.

Acknowledgments: the authors would like to thank the association of medical doctors 'ALS Associazione Liberi Specializzandi - fattore 2a' for facilitating this work.

Contributions: LT and SD contributed equally to this work, devised the project, the main conceptual ideas, and proof outline, drafted the initial version of the manuscript; CC, GL, and MM collected the sample; LT conducted the analysis; SM provided the theoretical formalism for a statistical account of the sample; PP verified the analytical methods and supervised the findings of this work. All authors have seen and approved the final version of the manuscript.

Conflict of interests: the authors declare no potential conflict of interests.

Data availability: data supporting the findings of this study are available from the corresponding author, LT, upon reasonable request.

See online Appendix for Supplementary materials.

Received for publication: 26 January 2021.

Accepted for publication: 9 February 2021.

This work is licensed under a Creative Commons Attribution NonCommercial 4.0 License (CC BY-NC 4.0).

©Copyright: the Author(s), 2021

Licensee PAGEPress, Italy

Italian Journal of Medicine 2021; 15:99-106

doi:10.4081/ijm.2021.1470

Introduction

The Italian Medical Residency test

Medical residencies in Italy are attributed to graduates from medical schools after a national test. According to the ranking, each candidate might then choose its training program until the fulfillment of positions. In Europe, centralized and national residency tests are not uncommon. For instance, a knowledge-based test adjusted for academic merits has been the adopted method for matching medical graduates in France, Spain, and Portugal. As the labor mobility between countries is one of the tenants of the EU project, several treaties have been enacted to regulate the reciprocal recognition of medical degrees and specialty titles.¹ However, the organization and definition of post-graduate programs still fall under national jurisdiction. While the test is defined at the national level, a high degree of autonomy is given to each of the forty-two Italian Medical Universities in defining courses and exams for their students. No evaluation system has yet been adopted to evaluate the level of preparation each University should match.

The national Italian residency test (SSM) comprises 140 questions, granting 1 point each for a correct answer and -0.25 for mistakes. On top of the test results, individual ranking positions are adjusted with 7 additional points based on the outcome obtained in the University of origin and the individual merits. The factors determining curricula scores can be found in Table 1.

The current scenario

While the residency test is national, the University curriculum depends on the University of origin, and it shows a high grade of autonomy for what concerns Grade Average and Graduation Mark scores. Even though curriculum scores give only 4.7% (7 out of 147) of the final score, this small percentage is crucial for selecting the candidates in the SSM. Italian SSM is highly competitive, as only a considerably reduced fraction of the candidates will obtain a residency position each year. In 2019, only 52.3% of candidates were matched to a residency position. In total, 17,565 candidates competed for 9183 residencies, 8075 of which were indiscriminately available to all of them (46.0% of the candidates) in the form of national grants, publicly funded.² Another important factor to consider is that the candidates with the highest final SSM score will have priority in choosing both the city in which to conduct the residency and the specialization. The available slots will progressively reduce so that candidates may even withdraw as they do not see

the available choices as fit.² This evidence may also partially explain why many doctors decide to seek a position abroad right after their graduation.³ In fact, from 2005 to 2015, 10,104 Doctors in Medicine from Italy applied to the UE scheme for professionals moving abroad and left the country.³ Also noted, 52% of all European doctors who moved abroad through similar schemes were from Italy.⁴ Although impossible to differentiate between specialized and unspecialized doctors within these numbers, the magnitude of the phenomenon is deemed relevant for the discussion. No study has previously attempted to validate this construct of interest to the authors' knowledge.

Validity framework

As the study aims to evaluate an established fashion in assessing education for medical graduates, an overall judgment is needed on the degree of the adequacy and appropriateness from which actions and interpretations about scores derive.⁵ This includes the current use of the score, which is under study and both implicit and explicit meanings attributed to it.

The explicit meaning attributed to SSM exams is that a better result directly translates into a better candidate, irrespective of the specific specialties under evaluation, as the exam is one and does not differentiate based on results in sub-fields of medicine. The implicit meaning of the curriculum scores is that the grading system used throughout Italy, in different

Table 1. Composition of Curriculum scores.

| Grade average | Points |
|----------------------|--------|
| Between 27 and 27.50 | 0.5 |
| Between 27.50 and 28 | 1 |
| Between 28 and 28.50 | 1.5 |
| Between 28.5 and 29 | 2 |
| Between 29 and 29.50 | 2.5 |
| Above 29.50 | 3 |

Note: grades range from 0 to 30, 18 being the minimum passing score.

| Graduation mark | Points |
|---------------------|--------|
| Between 105 and 108 | 0.5 |
| Between 108 and 110 | 1 |
| 110 | 1.5 |
| 110 and honors | 2 |

Note: grades range from 0 to 110, 66 being the minimum passing score.

| Other qualifications | Points |
|----------------------|--------|
| Experimental Thesis | 0.5 |
| Ph.D. | 1.5 |

Note: virtually all Italian Medical Graduates defended an experimental thesis on their graduation session.

medical schools, is either unanimous or that, although different, does not introduce bias. Another implicit meaning attributed to these curriculum scores is that higher grades should correct for potential variance in results through different years so that a candidate with good knowledge and marks might correct with additional points a result taken in a single point in time.

To the authors' knowledge, no literature was found on the correlation between the level of preparation of a single medical school and its impact on the alumni for future matching into residency programs. No literature, study, or report was also found about the correlation between the different levels of preparation offered by single medical schools in Italy and what SSM exams are concerned. Furthermore, for what concerns the specific case of the Italian residency exams, no literature, study, or report was found on the potential bias introduced by the different grading systems adopted by the single autonomous medical schools and the curriculum score attributed nationally irrespective of the University of graduation. The significant gap in literature reflects the absence of an overall system of evaluation for medical schools and medical residency programs both in Italy and in Europe but also lacking in several countries worldwide. The absence of a standardized evaluation system could result in potential bias in selection criteria for residency programs, introducing marginal benefits to some candidates that are not unsupported by either merit or evidence.

We conducted correlation analyses between the test score and the grades given to the same candidate during its six-year graduate program in medical school for validity evidence. This was calculated based on the final grade at the time of graduation, and the intermediate examination marks taken as a single average. The data were examined first for the national test of 2019, then together with results from 2018 and 2017. The SSM is a medical reasoning exam that uses specific questions about clinical scenarios and more granular and specific knowledge of biochemistry, pharmacology, and basic sciences.

Aims of the study

As previously mentioned, if on the one hand, the SSM is national, Italian Medical Schools are numerous and heterogeneous. This consideration leaves open the question of whether a possible between-institutions divergency may have an impact on: i) the curriculum score, as there may be different criteria in the evaluations within each institution; and ii) the SSM score, as the test questions may be compatible with the background given by each institution.

In this study, we considered the implicit framework used to allocate human capital to medical residency programs in Italy. The current planning involves a centralized test, corrected by marks and degree average. In this

context, the primary aim of this study was to investigate whether differences in grading between the Universities could introduce a selection bias in the candidates due to discrepancies in evaluating and administering medical programs in a system of high local autonomy. We hence evaluated: i) the association between the curriculum score and the residency test score; ii) changes of this association across the different Italian Universities, hypothesizing the presence of at least a modest/moderate degree of heterogeneity.

Materials and Methods

Sample

SSM results are public but semi-anonymized to protect the privacy of candidates. We were able to attribute specific variables to the 2019 cohort and associated results. Test results quote a reference for each candidate for his/her identification entry to the national Medical Board. We thus obtained the following variables to each entry: birth date, gender, test score, curriculum score, University from which the doctor graduated, date of graduation, provincial Medical Board of registration (which often coincides with family residency).

Statistical descriptive scores for the 2019 sample can be found in Supplementary Table S1a. The number of candidates per University for the 2019 sample is reported in Supplementary Table S1b. For within-University levels, test results and curriculum scores were averaged by candidates that graduated from the same University.

Procedures

The statistical packages R 3.6.1 and RStudio 1.3.959 were used in order to analyze data.⁶ The following libraries for R were used: *tidyr*, *ggplot2*, *ggridges*, *ggthemes*, *ggrepel*. The distribution of test and curriculum scores were evaluated through measures of central distributions (mean, median, mode), spread (variance, maximum and minimum), and percentile values. Results were graphically illustrated as density plots, with a vertical line to show the mean. Measures of correlation between curriculum and test scores were measured through Spearman's correlation coefficients. When correlation coefficients were calculated by University, NA values were discarded, as well as the only candidate from *Roma Università Pontificia Salesiana* as it was not possible to evaluate correlation coefficients by a single observation. Candidates graduated from abroad were excluded from all the analysis apart from Results 3.1 and the associated Control Analyses 3.3.1 & 3.3.4.

Spearman's rho correlation coefficients between test and curriculum scores were calculated, including

the 2018 and 2017 years, both for the national and local analyses. Rho correlation coefficients were used to evaluate the degree of stability of the correlations between Test and Curriculum scores across the three considered years was measured. The confounding effect of a longer time since graduation was evaluated through β coefficients, controlling for the potential bias of a longer time to prepare for the test.

Results

Validity and data distribution

A moderate/high correlation between Curriculum scores and Test results was found when considering the national sample in 2019 (Spearman rho=0.606; P-value<2.2e-16). Curriculum scores ≥ 6 can be seen to follow a different trend, which is comprehensible if it is considered that the sample size in this extreme upper bucket is limited. See Figure 1 for results.

The distribution of data was visually assessed through density plots. Curriculum scores showed a bimodal distribution, whereas test results were normal

although negatively skewed. The density plots can be found in the Supplementary Materials as Figure S1.

Validity by University

We graphically illustrated Test and Curriculum scores on a scatter plot, averaging results by University. The results can be found in Figure 2, where mean values for the x- and y-axis were reported using red dashed lines. Homogenous distribution of Curriculum scores with the Test scores would have been graphically represented by a linear distribution of the different Universities, as a lower curriculum would have corresponded to a lower grade in the test. It is possible to appreciate how Universities in the upper-left and lower-right quadrant reported discordant correlations between the two scores evaluated.

When the correlation was calculated intra-universities, coefficients showed a moderate degree of variability - spanning from 0.899 to 0.447. See Figure 2 for results and Table 2 for a detailed account of the correlation coefficient, together with 2018 and 2017 results. Distribution curves for test and curriculum scores were also plotted for each University;

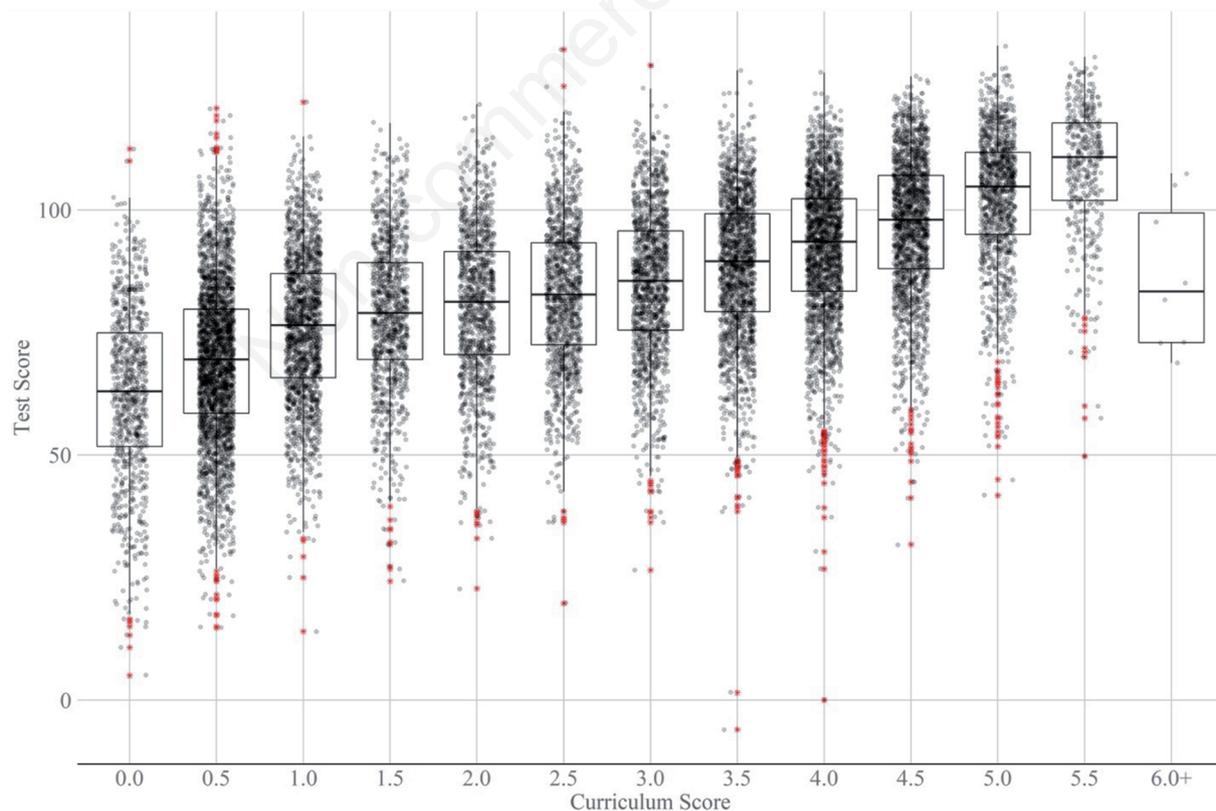


Figure 1. Correlations between Curriculum and Test scores. Boxplots represent the overall distribution for each Curriculum score class (from 0 to ≥ 6 , by 0.5). Individual variables were plotted as a jitter plot and outliers marked by the red color and a star shape.

they can be found in the Supplementary Materials as Figure S2.

Control analyses

Validity

We repeated the measurement of Spearman's correlation coefficients between test and curriculum scores, both at the national and local level, including 2018 and 2017 results. The correlation between the two scores at the national level remained stable (Spearman $\rho=0.600$; $P\text{-value}<2.2e-16$). The different trend of curriculum scores ≥ 6 was conserved, for which the reduced sample size must be retaken into consideration (see Figure S3).

Validity by University

Similarly, the correlation by University of graduation saw local correlation coefficients remaining relatively stable. Most Universities reported stationary coefficients of correlation through time, in different years. The correlation reached in 2019 by *Humanitas University* ($=0.899$) was calculated over 6 candidates only. *Humanitas University* can be considered an exception being poorly consistent with the same univer-

sity's correlation scores in the previous years: it was hence excluded as an outlier for the discussion. Table 2 reports correlation coefficients, per University, per Residency Test (2017, 2018, and 2019), together with the standard deviation between the three observations. Supplementary Figure S4 reports the Spearman's rho coefficients in the form of a heatmap.

Stability through time

As the standard deviation between the three measurements varied between Universities, a second heatmap was also graphically illustrated to have visual feedback on the stability of the results through time. Results can be found in Supplementary Figure S5. Excluding the *Humanitas University*, the highest SD was 0.092, which indicates a low variability between the different years of the residency test by the single institutions.

Days passed since graduation

As not all the students had the same time to prepare for the Residency test, the confounding effect of time passed between the graduation day, and the test was considered. A linear regression analysis was performed, using the days passed between the two days to predict

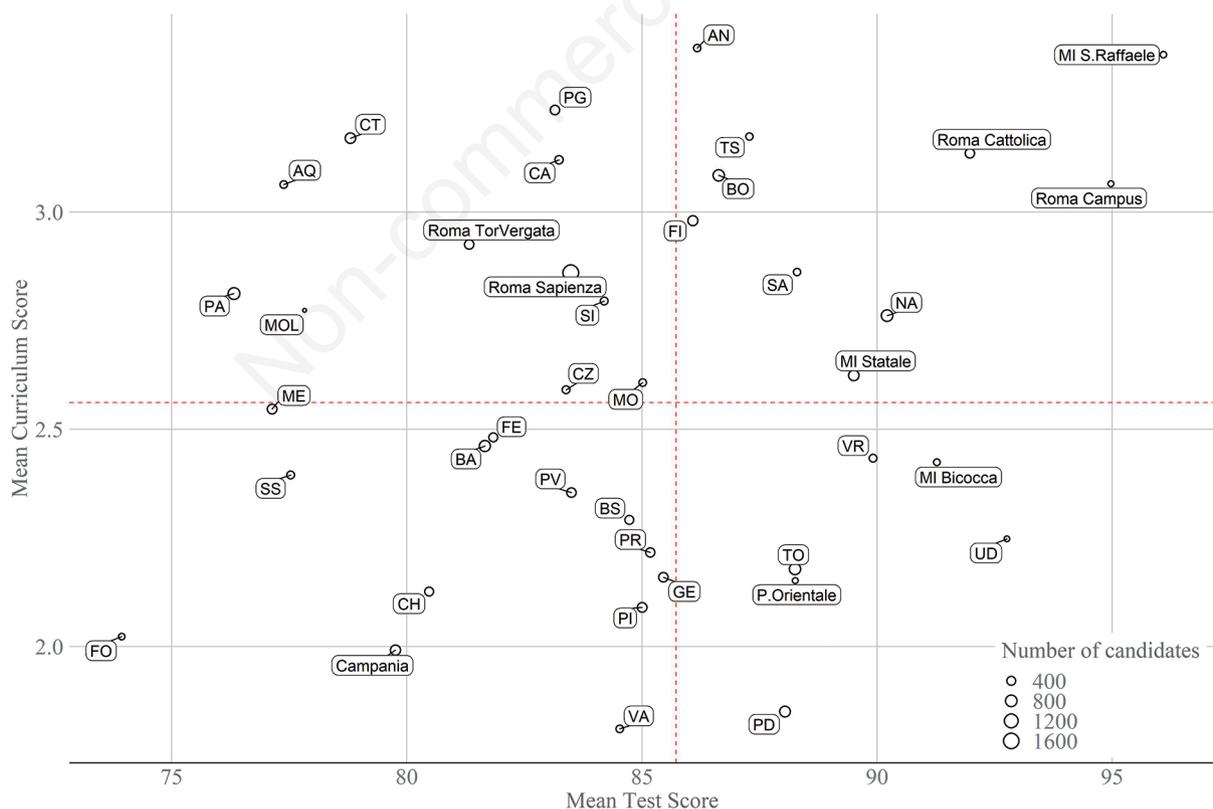


Figure 2. Correlations between Curriculum and Test scores - at the university level. Red axes represent the mean values over the sample. *Humanitas University* and *Roma Università Pontificia Salesiana* were excluded due to the low number of students participating in the test.

Table 2. Spearman's rho coefficients between test results and curriculum scores.

| | SSM2019 | SSM2018 | SSM2017 | S.D. |
|---|---------|---------|---------|-------|
| Humanitas University | 0.899 | 0.220 | 0.507 | 0.341 |
| Roma Università Cattolica Sacro Cuore | 0.788 | 0.720 | 0.641 | 0.074 |
| Milano-Bicocca | 0.775 | 0.742 | 0.648 | 0.066 |
| Università Degli Studi Del Piemonte Orientale Amedeo Avogadro | 0.742 | 0.796 | 0.674 | 0.061 |
| Verona | 0.731 | 0.654 | 0.561 | 0.085 |
| Roma Campus Bio-Medico | 0.730 | 0.711 | 0.608 | 0.066 |
| Milano Statale | 0.722 | 0.683 | 0.705 | 0.020 |
| Sassari | 0.721 | 0.695 | 0.550 | 0.092 |
| Milano Vita-Salute San Raffaele | 0.707 | 0.699 | 0.685 | 0.011 |
| Trieste | 0.705 | 0.706 | 0.648 | 0.033 |
| Siena | 0.697 | 0.735 | 0.606 | 0.066 |
| Firenze | 0.696 | 0.730 | 0.677 | 0.027 |
| Genova | 0.692 | 0.680 | 0.647 | 0.024 |
| Udine | 0.690 | 0.688 | 0.670 | 0.011 |
| Palermo | 0.673 | 0.630 | 0.630 | 0.025 |
| Modena e Reggio Emilia | 0.672 | 0.708 | 0.662 | 0.024 |
| Parma | 0.668 | 0.636 | 0.578 | 0.045 |
| Università Degli Studi Del Molise | 0.661 | 0.711 | 0.671 | 0.027 |
| Torino | 0.659 | 0.680 | 0.700 | 0.020 |
| Bologna | 0.659 | 0.695 | 0.622 | 0.037 |
| Padova | 0.657 | 0.607 | 0.608 | 0.029 |
| Roma Tor Vergata | 0.654 | 0.599 | 0.557 | 0.049 |
| Roma La Sapienza | 0.654 | 0.677 | 0.628 | 0.025 |
| Pisa | 0.646 | 0.659 | 0.641 | 0.009 |
| Ancona - Politecnica Delle Marche | 0.644 | 0.642 | 0.596 | 0.027 |
| Catania | 0.638 | 0.683 | 0.617 | 0.034 |
| Ferrara | 0.637 | 0.669 | 0.628 | 0.022 |
| Cagliari | 0.629 | 0.519 | 0.596 | 0.056 |
| Pavia | 0.624 | 0.624 | 0.662 | 0.022 |
| Perugia | 0.622 | 0.693 | 0.657 | 0.036 |
| Brescia | 0.620 | 0.689 | 0.704 | 0.045 |
| Napoli Federico II | 0.593 | 0.641 | 0.575 | 0.034 |
| Catanzaro | 0.590 | 0.627 | 0.569 | 0.029 |
| Bari | 0.553 | 0.601 | 0.554 | 0.027 |
| Messina | 0.550 | 0.603 | 0.569 | 0.027 |
| Chieti Gabriele D'Annunzio | 0.549 | 0.611 | 0.576 | 0.031 |
| Varese Università Dell'Insubria | 0.546 | 0.453 | 0.603 | 0.075 |
| Università Degli Studi Della Campania Luigi Vanvitelli | 0.544 | 0.619 | 0.652 | 0.055 |
| Salerno | 0.541 | 0.594 | 0.569 | 0.026 |
| Foggia | 0.478 | 0.518 | 0.522 | 0.024 |
| L'Aquila | 0.447 | 0.570 | 0.529 | 0.062 |

S.D., standard deviation.

the test score of SSM2019. When the whole sample was considered, a significant and relatively neutral relationship between the two variables was found (estimated β -0.004 , P -value $<2e-16$). We then included only the students who graduated less than a year before the test, and the relationship became positive while still significant (estimated β 0.083 , P -value $<2e-16$). When we included the students who graduated less than two years before the test, the relationship became non-significant and fairly neutral again (estimated β 0.002 , P -value 0.06). Figure S6 graphically represents the linear regression for the sample of graduated students from less than 5 years before the Residency Test to SSM2019. Two separate linear regressions were applied, one for students who graduated less than a year before the test and one for the rest.

In conclusion, we determined that the confounding effect introduced by the time passed between the graduation and test, although significant, was minimal at the aggregated level when evaluating results between Universities. Further results can be found in the Supplementary Materials as Table S2. To be noted, 2.4% of the sample (426 out of 17,955) included doctors who graduated more than 10 years ago.

Discussion

This study identified a significant correlation between the curriculum score and the test score for all Italian Universities. The degree of correlation varied according to the University considered, but the correlations ranged from moderate ($r=0.447$) to high ($r=0.788$). The high majority of the universities showed correlation coefficients between 0.5 and 0.7, over 70% of the sample. These data suggest a modest but not negligible between-university variability, which can also be interpreted as a moderate level of consistency in both the preparation and evaluation systems adopted by single medical schools. We considered: i) differences in the test structure between 2017, 2018, and 2019; and ii) time passed since graduation as possible factors influencing such variability. Test-curriculum correlation scores were very stable within each University (highest $SD=0.092$): it is hence probable that the variance is attributable to the University of origin rather than to the heterogeneity of the test or class year by year. Time passed since graduation was an important factor, showing a positive influence in the first year after graduation (interpretable as a higher time available to prepare the test) and a negative influence after the first year (interpretable as a less updated knowledge coming from the preparation given by the University of origin, together with the inverse survival bias induced by candidates at successive attempts).

These data may be used as a starting point to favor the coordination of the single faculties with the national

SSM committee. For instance, the programs of courses that highly correlated with the test may be compared to the less correlated ones to better match the test with the backgrounds provided by each University.

Considering these factors, a thorough review of the residency test seems warranted for what concerns the long-term programming of the healthcare system.

Limitations

A plausible factor that may influence the final SSM score is participation in paid training courses that the candidate may independently choose to attend. However, we have no information concerning these data.

Working while preparing for the SSM test could also be a confounding factor, which would also reflect the relative socio-economic status of the candidate and could be potentially associated with lower participation in the paid training courses described above. Either socio-economic information about the candidates or their working status was collected, which may explain a portion of the observed variance in the results.

Even though the total test score is a valid parameter on which we could base the analyses we conducted, the score needed to pass the test varied considerably for each residency school (in 2019, if only nationally funded scholarships are considered, it ranged from a minimum score of 110 for Plastic Surgery to 82.25 for Microbiology & Virology - regional or private grants induce a secondary distortion in minimal scores).

Conclusions

Correlation differences between Test and Curriculum scores across institutions are important, as even minor variations of the final score may determine admission or not for a large number of candidates. The present work may help to define the criteria in the formulation of the national residency test questions. Our work also highlights the importance of coherence between the tests through different years. This allows each University to identify the specific contents to be stressed during the course, thus enhancing its students' preparation.

Finally, further research may focus on exploring similar systems in other countries. Assessing correlations between the local (at the end of the university course) and the national level (SSM test) would allow comparing systems on an international scale, progressively reaching shared criteria in the medical training and evaluation.

References

1. Directive 36/EC of the European Parliament and of the Council on the recognition of professional qualifications

- [Internet]; Sep 30, 2005. Available from: <http://data.europa.eu/eli/dir/2005/36/oj/eng>
2. Decree n. 859 from the Minister of Education for the admission test in the Italian medical specialty schools [Internet]; May 2, 2019. Available from: <https://www.universitaly.it/index.php/content/ssm1819-normativa>
 3. European Commission. Regulated professions database [Internet]; ongoing [cited 2020 Apr 16]. Available from: <https://ec.europa.eu/growth/tools-databases/regprof/index.cfm?action=homepage>
 4. Enpam - Eurispes. Il termometro della salute 1° Rapporto sul Sistema sanitario [Internet]; 2017 [cited 2020 Oct 15]. Available from: <https://eurispes.eu/ricerca-rapporto/enpam-eurispes-il-termometro-della-salute-1-rapporto-sul-sistema-sanitario-2017/>
 5. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Meas Issues Pract* 1995;14:5-8.
 6. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria; 2020. Available from: <https://www.R-project.org>

Non-commercial use only