

DOI: <https://doi.org/10.54937/2026.9788056112410.25-31>

ZAMYŠLENÍ NAD AKADEMICKÝMI HADY A ŽEBŘÍKY – JAKÁ JSOU PRAVIDLA HRY A JAKÉ MÁ TATO HRA DŮSLEDKY PRO HODNOCENÍ AKADEMICKÝCH PRACOVNÍKŮ?

ON THE ACADEMIC SNAKES AND LADDERS GAME – WHAT ARE THE RULES AND IMPLICATIONS FOR ACADEMIC STAFF EVALUATION?

JAN STOKLASA¹, JANA STOKLASOVÁ²

¹ LUT University, Lappeenranta, Finland / Yliopistonkatu 34, 53850 Lappeenranta, Finland; and
Palacký University Olomouc / Křížkovského 511/8, 779 00 Olomouc, Czech Republic
Email: jan.stoklasa@lut.fi, jan.stoklasa@upol.cz
ORCID iD: 0000-0002-1810-9181

² LUT University, Lappeenranta, Finland / Yliopistonkatu 34, 53850 Lappeenranta, Finland
Email: jana.stoklasova@student.lut.fi
ORCID iD: 0000-0002-0145-6523

ABSTRAKT

V tomto článku se zabýváme „akademickými žebříky“ a „akademickými hady“ – žebříčky a dalšími nástroji, které nám umožňují lepší orientaci v akademickém prostředí. Od těchto pořadníků, ať už univerzit, zemí, fakult, výzkumných týmů, jednotlivců, ale i studijních programů a celých vědních oborů, očekáváme, že nám poskytnou jednoduchou a srozumitelnou podporu rozhodování a orientaci, případně vzhled do poměrně složitého akademického prostředí. A v mnoha případech tento účel opravdu plní a efektivně komunikují relativní výkon jednotlivých porovnávaných entit (institucí, jedinců, ...). Dávají nám odpověď (přirozeně interpretačně závislou na použité hodnoticí metodologii na niž je následně založeno uspořádání jednotlivých entit) na otázku komu se aktuálně „daří lépe“ a mohou nám také pomoci identifikovat příklady dobré/špatné praxe. Jelikož jsou potenciální informační přínos a použitelnost žebříčků obecně akceptovány, tento článek si klade za cíl upozornit na potenciální cenu, kterou za jejich používání platíme – zejména v kontextu kontinuálního hodnocení a sledování vývoje. Snaží se tedy především klást otázky ohledně možných nedostatků a selhání žebříčků jako podpory rozhodování, a ztráty nebo zkreslení informací v nich obsažených. A to v kontextu hodnocení kvality a výkonu, s bližším zacílením na jejich použití v kontextu hodnocení akademických pracovníků.

Klíčové slová: hodnocení, akademická sféra, výkon, relativní, absolutní, žebříčky, agregace

ABSTRACT

This paper discusses the “academic ladders” and “academic snakes” – rankings and other various tools – that help us get oriented in the academic landscape. The rankings, be it of universities, states, schools, research teams and even individuals, as well as of study programmes and whole fields of study, are expected to provide a simple and understandable decision support. And in many cases, they actually

succeed in communicating the relative performance of the entities being compared. So, they do give us the answer (obviously based on the methodology that is being applied to evaluate the entities and then rank them based on the evaluations) concerning who is currently “doing better” and they might even help in the identification of the examples of good/bad practice. While the potential usefulness and benefits of the rankings seems to be clear, this paper aims to draw attention to the potential cost of the use of rankings, particularly in the context of continuous evaluation and development tracking. In other words, it is intended to ask question concerning the drawbacks, loss of information and other related issues of rankings, and the relative information contained therein, being applied in performance and quality assessment, with a more profound focus on the area of academic staff performance evaluation.

Keywords: *evaluation, academia, performance, relative, absolute, rankings, aggregation*

INTRODUCTION

In the academic world we have grown accustomed to the constant benchmarking, comparisons and ranking in almost every area of our activity. The universities are being ranked based on various criteria and rankings of universities a research institutions, education institutions (in general, or in specific fields and sub fields) are being provided to the world and everyone who has the need for such information (see e.g. [1-3] and many other providers of such rankings). These ranking lists and benchmarks are expected to provide a highly aggregated piece of information concerning the performance of a university or its unit or a scientific field covered by the university in comparison with the others. Well, the “comparison with the others”, that is the relative nature of the “ladder” based on the ranking is not being stressed that much. On the other hand, the “top 100”, “top 300-350” or “top 10” etc. positions on the ladders are easily used in advertising, in communication with the stakeholders of these institution and, in essence, in the assessment of the quality of these institutions by the general public, the applicants, the potential employees, the current employees, the research and or education partners, the funders and so on. After all, the information is provided on an easy to use ordinal scale, it is simply a ladder and the higher we managed to climb, the better we must be (given the applied evaluation methodology), correct?

The companies providing these rankings usually disclose the methodologies applied for the ranking and also admit their potential limitations. So the evaluation and the corresponding ranking (ladders) should be objective, correct? But where do the ladders actually stand in the strategy, process of goal achievements and in the realisation of the very purpose of the tertiary education institutions and their most values assets – the people working and studying there? Without any loss of generality (hopefully) the paper uses the metaphor (or model?) of the children’s game of snakes and ladders to make the general ideas clear to the general public including future students that are, after all, one of the key segments of recipients of the information represented by various rankings. Let us set the game and have a look at its rules first.

THE RULES OF THE GAME – ACADEMIC EDITION

If we oversimplify, then the goal of the game of snakes and ladders is simply to climb to the top – well not exactly, it is to get to the final field of the plan. The higher (further) we get on the plan, the closer we get to the overall goal of winning the game. There are ladders to assist us in getting up faster, and there are snakes pushing us down and in-between there is the continuous approach towards the final (upmost) field that might be slow, but unless we encounter a snake, it is monotonically approaching the final field. And the first one that reaches the final field wins. Also, the speed of our approach through the gameplan is to some extent influenced by a stochastic process, usually represented by a dice.

How close a model could this be in the academic environment? Let us take a closer look:

- The overall goal of getting (close) to the last field on the gameplan is relatable and the universities and the academics probably have a set goal they want to achieve. In the institution perspective this is embodied by the strategy and the corresponding goals, for the academics the goals can be promotion/research results/contribution/teaching-quality related. So there is a trajectory aiming towards the final goal. True, the game plan might not be fully known and it can be changing, but the overall trajectory upwards and towards the end field of the plan seems compatible with the game model.

- We move forward from the starting point continuously by gaining experience and expertise, by developing and by “growing”, with speed determined not only by our capacities, but also some random elements – that also might be compatible with the game model.
- There are ladders that help us advance further – in the academic world these can be represented by the rankings (almost actual “ladders”) of institutions, research fields, successful funding recipients or of highly cited researchers, highly impactful researchers, highly cited papers, numbers of patents, societal impact, stakeholder reception etc. Note that these ladders do not necessarily aim towards the end goal we have set, but climbing up can help us get where we need to get. It can provide us with more credibility, it can increase our attractiveness as an employer, it can increase our chances of getting talented students and faculty and thus producing high quality research/educational results faster etc. So the ladders seem to fit the model as well, and they also seem to be a tool to be used to get to where we want to be, but not necessarily directly aimed toward the set goal by design. However, it is not guaranteed that one would climb all the way up a specific ladder in the academic context, but just some way up.
- There are also snakes – these can be embodied by various setbacks, failures, mass paper retractions, data manipulation scandals, papermill activity involvement, other ethical issues, negative externalities, And these might also not be related directly to the overall goal, but can still affect the professional reputation and attractiveness of the institution or the individual, thus increasing the distance between them and the final goal. This also seems to be compatible.
- The first person/institution reaching the final field of the game plan wins – in this case it seems that there might be a slight difference between the model and the academic reality as the academic endeavour does not seem to be a race (at least not by choice/design). However, the universities are competing for a limited set of students and academics, and the further (higher) they are on the game plan, the better chances they might have to attract the most suitable people, who will in turn help maintain and increase the momentum of their journey through the plan. Also other resources we are competing for are limited and being further along the game plan can increase the share of the strategic resource one could strive for or obtain.
- The time of availability and the length of the ladders and snakes might also have a random component, which would be analogous to the game, but the universities and academics can actually actively try to increase the chances of encountering ladders by performing well (in this case relative to others¹) and decreasing the probability² of encountering the snakes by following scientific rigour, the ethical principles and by aligning their activities with the set goals (assuming the goals themselves are ethical and in line with good scientific and educational conduct).

In essence there are many similarities and the game itself could serve as a nice analogy (if not an actual model) of the “academic struggle for success”. It even seems that in the academic version of the game the players have a bit more control over the “tools” (ladders) and “traps” (snakes) – at least over their length, number and probability of occurrence.

ACADEMIC SNAKES AND LADDERS – INDIVIDUAL EDITION

If we now look away from the “team” version of the game (university ranking etc.) and have a look at the more individual game sessions, it seems that the level of control over the probabilities of snake occurrences is even higher, as many issues related with the snakes (and the connected fast or slow slides towards the beginning of the gameplan) are linked with conscious decisions of the teachers and researchers (academics). To oversimplify, if one does not try to rig the game in his/her favour by cheating (behaving unethically, taking advantage of the others, taking credit for someone else’s work, cutting corners in teaching, buying publications, papermill activity involvement, irrelevant or fabricated or incorrect results publications etc.) the snakes are much less likely to lie in one’s way toward the final field of the game plan. There can still be an occasional snake or two stemming from honest mistakes,

1 As rankings are by definition based on the relative position of the evaluated object in the full list on objects being considered. As such the information contained in them is relative (ordinal).

2 Or at least modify the probability density function of their appearance.

unpredictable or uncontrollable conditions etc. But there seems to be a high level of control over the length and number of the snakes that can be consciously exercised by the individual academics.

For the completeness of the presentation of the game metaphor, willing or not all the tertiary education institutions and all the academics are a part of some version of the game. Their role might be more active or more passive, their chosen speed of advancement towards the top (and towards the final field of the gameplan) can differ, but they (we) are all playing it. And with the limited resources it is at least partially a race to the top. And all the players involved are also actively changing the game plan³; also the length, number, placement and stability of the ladders (and snakes) is conditional on the whole academic context and all its players⁴. In the case of, for example, the highly cited researcher list [4], one does not even provide information and the ranking is being created anyway. It is not easily possible to opt out of the list, to opt out from providing context for the ranking etc. And since research and the publication of its results is a crucial part of the duties of academics, alongside the education of the new generation of scholars and practitioners, contribution to the development of one's field of science, to the enhancement of our understanding of the world etc., we cannot easily claim these rankings to be irrelevant or useless. They do provide some information. So some ladders are available to everyone, some are available to those that actively seek them. The number of available ladders might also be region-, field of science-, context-, time-dependent. And so might be the number of snakes... What is worth noting is the fact that the snakes and ladders (both in the individual and in the "team" version of the game) might not be completely independent. It might be that if specific choices/adjustments are made⁵, some ladders appear sooner, become higher and/or our ability to climb them is enhanced. However, optimizing our behaviour for one particular type of ladders can create more snakes and speed bumps in the other parts of the game plan. On the other hand, a careful avoidance of snakes can result in a compromised ability to find or climb ladders⁶.

Also note that while for established tertiary education institutions and fields of science it might be difficult to actually lose the game, it can still happen. It is, however, highly unlikely (but not completely impossible) to be disqualified from the game. However, in the individual edition, one can actually "take a snake slide under the table", that is, one can be disqualified, or his/her chances of ever playing the game again can be limited or the speed of advancement through the game plan can be significantly reduced by not having any ladders to use or only having very short ones.

This seems to imply that in the broad context of moving through the game plan towards the final field and towards the overall goal, the ladders (and snakes alike) are mainly tools that can speed up, change direction or simply modify the trajectory of the player. As such, they are not the most important part of the game. They do influence the game dynamics, they can favour some players more than others (mainly those for whom there are more ladders and less snakes to encounter or use), but they are not crucial to the actual completion of the game. One might not be among the fastest to reach or approach the final field of the game plan without ladders, but it is definitely possible to get there.

3 The overall goals change in time, the strategies react to that, new problems to be solved emerge, some problems get sufficient solutions. The academia is simply a living ecosystem that has its purpose and evolves in time alongside the society it is supposed to serve.

4 For example, in university rankings (see e.g. [1-3]), it is clear that only the universities actively participating and providing the information are being ranked. As such the "top 100" does not mean 100 of the absolutely best, but the best hundred of those being ranked (given the criteria and evaluation method that was applied)). This means that those that are not ranked are "not visible" and as such cannot be seen as good/strong based on the ranking, but they cannot also appear at the bottom of the ranking.

5 It might not necessarily be cheating or unethical conduct – one could use the term "optimization" as a fitting term. People, including those in the academia, tend to modify and adapt (optimize) their behaviour towards more efficient one. This is not necessarily a bad thing, but system-wise, if the optimization is too massive and goes against the purpose of the system, it can be dangerous.

6 For example, if one does not publish, one does not risk any retractions or research integrity issues. However, this implies nothing to be cited and as such no citation-based ladders to ever be available.

Now that we have an idea about the game that is being played and about its rules and the level of control that we as individual academics have over the gameboard and its elements (snakes, ladders) it might be a good time to start asking additional question. And since we are playing the game on a professional level it might be good to be able to identify examples of good/bad practice, “good” players, “cheaters”, “average” players”, players that need more training, coaching, resources, or assistance to be able to play better or to get further on the game plan.

ASSESSING THE PERFORMANCE OF INDIVIDUAL PLAYERS OF THE ACADEMIC SNAKES AND LADDERS

In the context of the academic game(s) that we play, it might be interesting (and in many cases very important) to know who is playing well and who’s performance needs adjustment or who needs support to be able to play better. This is where the evaluation fits in. Within the snakes and ladders metaphor, it is not necessarily important who got to the final field the fastest or who got closer than the others. Mainly if we have set the game such that every individual defines (with some externally set constraints) his/her own content of the final goal. As such the proximity to the final goal does not carry much information if the final goals are significantly different. However, this is the very usual case in the academia, where different institutions, different fields of science/education, different sub-domains of these fields, different academic ranks etc. are expected to be performing differently.

If this does not complicate the use of rankings in decision support and evaluation enough, the ordinal (and relative) nature of the rankings complicates things even further. As rankings are by definition ordinal, their interpretability and understandability is conditional on having a clear and easily identifiable prototype of good performance (and bad performance) in the set of objects being ranked. If such a benchmark is not established, there is no guarantee that the “top ranking” object is good enough, or that the “bottom ranking” object is bad. And this is even more problematic in the context of rank reversal issues connected with relative-type evaluation methods. The rank reversal problem is, simply put, a feature of an evaluation method that potentially results in a change in pairwise ranking of objects if some other objects are included in or removed from the set of objects being ranked. In other words, it means the dependence of the pairwise rankings of objects on the composition of the set of objects being compared (see e.g. [5-7]). Note that if the set of ranked objects can change (which is the case of the university rankings as well as of the academic staff evaluation within a specific university or its unit) the pairwise rankings can potentially change as well if a relative-type evaluation method is used. Relative-type evaluation methods are all those multiple-criteria evaluation methods that derive the value of objects from their comparisons with other evaluated objects (either directly, or in standardization of the values of criteria or in other steps/calculations).

This would suggest that while potentially simple to interpret and understand, rankings might not be the most valuable source of information for the management and support of academics. It might be more important to understand how far the academics are from the usual (expected/required) and good performance (in terms of composition, quality, quantity of activities and their impact), where do they struggle, where do they perform well. We do not necessarily need to know who performs better and how are academics ranked in terms of their performance for several reasons:

1. because the relative-type information without a clear prototype (or definition) of “good performance” and “bad performance” might be difficult to interpret
2. because the ranking does not actually allow us to manage the most valuable resource we have any better
3. because even within a single department or research team the courses being taught, students being supervised, papers being published, topics being researched can differ so much that a direct comparability might be difficult
4. because all the evaluation models are, by definition, qualified simplifications of reality and as such they leave some aspects out – automatic ranking might thus not even be objective or meaningful
5. because it might be more important to know whether the unit works sufficiently and whether its goals and the goals of the upper-level units in the hierarchy are being achieved
6. because ranking cannot be planned and is quite unpredictable for the individuals being evaluated

Even though there might be other reasons while not to aim for rankings in staff evaluation in general, the sixth point might deserve a bit more attention here. The main issue connected with this point is that the ranking of an individual is not dependent on his/her own performance only, but also on the performance of the others. As such it is possible that if an individual performs in an identical way this year and the next one, he/she will be ranked differently in both years. This also means that a clear improvement in the performance of an individual can still result in his/her worse ranking (assuming that others have improved their performance more with respect to the previous year), or that performing worse than in the previous evaluation period can potentially still result in a better ranking than in the previous period (assuming that the performance of the others got worse relatively more). As such, it is meaningless to set ranking goals unless we can be certain that everyone has equal conditions and resources to perform their tasks. It is also potentially demotivating for the academics (well for any workers) to be clearly better than last year and yet register as “worse” (ranking-wise). It is true that the ranking along with its “relative position” information can provide additional insights, but it is difficult to be used as the main basis for evaluation. One more reason to avoid rankings as final evaluations is their high incompatibility with timeseries analysis – the ranking, being ordinal (relative) – does not carry the information concerning actual improvement or deterioration of quality or quantity of tasks, making the tracking of performance development highly tedious and impractical.

The evaluation systems can be designed also in such a way that avoids rank reversal completely, that provides evaluations that do not lose interpretability when their development in time is being analysed and that do not allow for the creation of rankings by design (at least not in the full evaluation universe, but potentially allow for rankings in those areas of the evaluation universe where additional resources might need to be distributed to improve performance or to level the playing field) – see e.g. [8-11].

CONCLUSION

Going back to the gameboard of snakes and ladders, in light of the unpredictability of rankings it is becoming clear that the reliance on ladders to speed up the process of advancement through the gameboard toward the final field can be risky and potentially unpredictable, as rankings are never fully under the control of the individual (or individual institution) being ranked. To improve ranking one needs to be able to improve the relevant performance criteria more than the others being ranked. To optimize this resource-wise it requires the ability to predict how the others will perform and also who will be ranked (as inclusions of new individuals also potentially modifies the ranking). It thus seems potentially inefficient and risky to rely on the ladders as the main means for the advancement through the gameplan. It might be more prudent to be able to advance through the game plan in a steady and professional way, with constant (or increasing) speed and confidence, gaining expertise, knowledge and reputation on the way. And then the snakes should not be too frequent, and an occasional ladder can create a nice and well deserved shortcut upwards. Even though seemingly informative, rankings constitute a potentially problematic piece of information for further analysis and for clear interpretability. Rankings need to be interpreted in the context of all the objects being ranked and ideally in connection with the underlying values leading to the rankings. They are also seemingly unpredictable as any direction of development of performance can lead to any direction of the change of the ranking. As such setting ranking goals, while theoretically possible, does not seem to be reasonable.

And where do we continue from here? As the rules of the game have been set (metaphorically speaking), we can now dive deeper into the definitions of different performance classes of individual and institutional players, identify what makes snakes appear more frequently, check whether there is a category of “snake charmers” who know how to behave around the snakes and as such do not suffer from their presence. Whether there is a class of ladder builders, good ladder climbers, players afraid to use ladders (and hence avoiding them, whenever possible), players over-relying on the ladders – and most importantly how does their performance in the game differ. Another interesting research direction that can be further pursued is the connection between the individual and institutional editions of the academic snakes and ladders game. Is there success spillover? If it is present, is it unidirectional or bi-directional? Is the effect of snakes/ladders strong enough so that a snake/ladder in the individual version can affect the gameplan

of the institutional version as well and vice versa? Luckily, the game is quite transparent and the moves of the players are rather well documented, so the required analytics should be well possible.

LIST OF REFERENCES

- [1] SHANGHAIRANKING CONSULTANCY. (2025). Academic ranking of world Universities. >https://www.shanghairanking.com/_pub/com/arwu/2025/2025%20ACADEMIC%20RANKING%20OF%20WORLD%20UNIVERSITIES.pdf<
- [2] QUACQUARELLI SYMONDS LIMITED. (2025). QS World University Rankings. ><https://www.topuniversities.com/university-rankings>< [accessed 28.12.2025]
- [3] TIMES HIGHER EDUCATION. (2026). World University Rankings 2026. ><https://www.timeshighereducation.com/world-university-rankings/latest/world-ranking>< [accessed 2.1.2026]
- [4] CLARIVATE. (2026). Highly cited researchers 2025. > <https://clarivate.com/highly-cited-researchers/>< [accessed 2.1.2026]
- [5] AMJAD, Hassan, & STOKLASA, Jan (2025). A commentary on the paper “Root Assessment Method (RAM): A novel multi-criteria decision making method and its applications in sustainability challenges” – Investigating the properties and identifying sources of rank reversal. *Journal of Cleaner Production*, 521(February), 146035. <https://doi.org/10.1016/j.jclepro.2025.146035>
- [6] AMJAD, Hassan, & STOKLASA, Jan (2025). The nature of the evaluation method matters. On the importance of “relative – vs. absolute-type” method choice in assessing the effect of tourism on UNESCO natural heritage sites. In M. Müller (Ed.), *Knowledge on Economics and Management, Conference Proceedings*. December 4-5, 2025, Olomouc, Czech Republic (KNOWCON 2025) (pp. 5–11). Palacký University Olomouc.
- [7] KREJČÍ, Jana, & STOKLASA, Jan (2018). Aggregation in the analytic hierarchy process: Why weighted geometric mean should be used instead of weighted arithmetic mean. *Expert Systems with Applications*, 114, 97–106. <https://doi.org/10.1016/j.eswa.2018.06.060>
- [8] TALÁŠEK, Tomáš, STOKLASOVÁ, Jana, STOKLASA, Jan, & TALAŠOVÁ, Jana (2021). Visualising HR data concerning the performance of university staff – career development assessment and assistance perspective. In R. Hlavatý (Ed.), *Proceedings of the 39th International Conference on Mathematical Methods in Economics* (pp. 481–486). Czech University of Life Sciences Prague.
- [9] TALAŠOVÁ, Jana, STOKLASA, Jan, HOLEČEK, Pavel, & TALÁŠEK, Tomáš (2017). Mathematical support for human resource management at universities. In P. Pražák (Ed.), *Proceedings of the 35th International Conference on Mathematical Methods in Economics* (pp. 783–788). University of Hradec Králové.
- [10] COLLAN, Mikael, STOKLASA, Jan, & TALAŠOVÁ, Jana (2014). On Academic Faculty Evaluation Systems: More than just Simple Benchmarking. *International Journal of Process Management and Benchmarking*, 4 (4), pp. 437–455. <https://doi.org/10.1504/IJPMB.2014.065522>
- [11] COLLAN, Mikael, STOKLASA, Jan, & TALAŠOVÁ, Jana (2016). Staff Evaluation Systems-Shaping Autonomy through Stakeholders.pdf. In R. Turcan, J. Reilly, L. Bugaian, eds. (Re)Discovering University Autonomy – The Global Market Paradox of Stakeholder and Educational Values in Higher Education. Hampshire: Palgrave Macmillan US, 2016, pp. 97–106.