



The mismeasure of higher education? The corrosive effect of university rankings

David Robinson*

Senior Advisor, Education International, 5 blvd du Roi Albert II, 1210 Brussels, Belgium

ABSTRACT: This paper examines the limitations and biases of world university rankings and asks what drivers explain their ongoing proliferation and popularity. It is argued that rankings are having a corrosive effect on higher education systems, institutions and staff by encouraging policy reforms at the governmental level and a reallocation of resources at the institutional level that may improve standings in the rankings but do not necessarily enhance quality research and teaching. Global rankings are linked to the rise of an international market in higher education, particularly with respect to international students. The author argues that what is at stake in the debate over university rankings is fundamentally whether higher education is to be thought of as having intrinsic value, or whether it is defined narrowly in instrumentalist and consumerist terms.

KEY WORDS: University rankings · Quality assurance · Assessment of higher education · Learning outcomes · Performance indicators

— Resale or republication not permitted without written consent of the publisher —

Introduction

In *The Mismeasure of Man*, noted American paleontologist and evolutionary biologist Stephen Jay Gould presented a devastating critique of the methods and motivations underlying biological determinism in the assessment of intellectual ability (Gould 1981). Gould exposed the erroneous attempts throughout history to measure the complexity of human intelligence, from the early crude experiments claiming empirical links between intellect and skull and brain size, through to more recent and ostensibly sophisticated quantitative tests which purportedly measure 'intelligence' as a single number or quotient for each individual. For Gould, each of these allegedly scientific and neutral efforts to rank people in a single series of intellectual worthiness was not only methodologically flawed, but also biased. Results were invariably used to show that specific disadvantaged groups—races, classes, or sexes—are innately inferior and, by extension, thus deserving of their status.

In reflecting upon the debate over university rankings today, one may well ask if a similar critique could be leveled against attempts to measure and order the worth or quality of higher education institutions. That is, to what degree do university rankings suffer from the same basic fallacies as 'scientific' measurements of intelligence? Do the rankers, like the biological determinists, share the mistake of trying to convert an abstract and complex concept into a single quantifiable property placed on a gradual ascending scale? Do university rankings have the effect of reproducing and justifying an existing hierarchy of institutions?

These questions cut to the core of what probably represents the most pervasive concern with university rankings—that what makes a good university simply cannot be accurately captured statistically. University rankings serve to reduce the complexity of an institution's achievements and practices to a simple ordinal scale that cannot in any real sense claim to be a valid indicator of quality. The quality of the education and free inquiry that takes place within an

institution cannot be easily or accurately parsed, quantified, ordered and compared. Quality higher education is not a singular product or outcome subject to one simple definition or numerical score. It has to do with a diverse range of activities and processes. Rankings require that these complex aspects of a university be reduced to a number no matter how absurd the exercise becomes.

The limits of ranking

What most rankings today arguably measure is not quality but more accurately the wealth of a college or university. This is reflected in some of the most common measurement criteria used: the value of a university's endowments and operating budget; its reputation as judged by the entrance grades of students and the opinions of leaders from business and the academy; the research awards and output of faculty; and student satisfaction as measured by the share of alumni donations. As a result, as Hazelkorn (2009a) observes, the universities that appear at or near the top of the major rankings are 'distinguished by large budgets, large endowments, age, excellent staff to student ratios, and most importantly, access to large pools of highly developed human capital (staff and students).'

Other measurement criteria have also been identified as problematic. Some of the popular international rankings count reputational factors that are at best subjective and at worst simply recycle existing biases and perceptions about a university's status (Guarino et al. 2005). One study, for instance, found that participants in reputational surveys were unfamiliar with as much as one-third of the programs they were asked to rate (Brooks 2005, p. 7). Some universities carry a weight of historical prestige and those taking part in reputational surveys may rate them higher based upon their perceived status rather than any real measure of their quality today. The reliance on reputational surveys therefore opens the question of whether rankings simply reproduce preconceived perceptions of quality and status.

Rankings also create the illusion that there is a clear and identifiable distinction between universities, when in fact the differences are often statistically insignificant — even if one accepts the questionable weighting of different categories. The 'league tables' produced are highly simplistic snapshots of a university's summative score, and open to manipulation. Indeed, there is good reason to maintain a healthy dose of skepticism when it comes to reading

the relative position of universities in a ranking. As management expert Henry Mintzberg suggests:

Anyone who has ever produced a quantitative measure — whether a reject count in a factory as a surrogate for product quality, a publication count in university as a surrogate for research performance, of estimates of costs and benefits in a capital budgeting exercise — knows just how much distortion is possible, intentional as well as unintentional. (cited in Birnbaum 2001, p. 79)

Almost all current rankings overemphasize research output and citations as a proxy for 'quality'. The number of articles faculty members produce is, however, not necessarily an indicator of the quality or impact of publications. As Altbach (2006a, 2010) argues, publication counts are most often drawn from established refereed publications included in databases such as the Institute for Scientific Information (ISI). The problem is that these databases contain mainly English-language journals and thus tend to reinforce the standing of large English-speaking universities in the United States and the United Kingdom (Altbach 2006b). In addition, universities with large medical schools and natural science departments tend to perform better because the nature of research in these disciplines is such that academics publish more articles compared to their counterparts in the social sciences and humanities.

Teaching: The missing variable?

Rankings also largely ignore one of the major components of what faculty do — teaching. This lacuna is partly explained by the fact that the quality and impact of teaching is much harder to measure statistically than research productivity. Counting up citations and patents awarded is one matter, measuring the impact of teaching is quite another. To be fair, some rankings such as the Times Higher Education World University Rankings and the QS World University Rankings have recently sought to assign certain proxies to measure teaching quality. These rankings now include reputational questions, student satisfaction surveys, student–teacher ratios, and the number of PhDs held per staff member. However, these measures are at best stand-ins for teaching quality and do not provide an accurate picture of what actually happens in the lecture hall.

While criticisms of rankings have been well rehearsed, they have prompted not so much a re-examination of the motivations of rankings as a tweaking and re-tooling of existing instruments. The failure of most rankings to factor in teaching, for

instance, has led to calls for new tools to assess teaching effectiveness and learning outcomes. The Organization for Economic Cooperation and Development (OECD) has taken up this cause, launching a controversial multi-million euro feasibility study aimed at developing the Assessment of Higher Education Learning Outcomes (AHELO) (Tremblay et al. 2012). The study is built upon 3 different tools or 'strands': a generic strand, based upon the Collegiate Learning Assessment (CLA) administered in the United States of America, that seeks to evaluate the skills that all students, regardless of discipline, should possess toward the end of their undergraduate degree (e.g. critical thinking, problem solving, and written communications); a discipline-specific strand that focuses on assessing the knowledge and abilities of students in engineering and economics; and a contextual strand that seeks to gather information about the institutional environment and background of students (OECD 2010). The objective is to use AHELO to quantify the 'value-added' learning provided by institutions across a number of countries both within and beyond the OECD. According to the OECD Secretariat, such a measurement tool 'could provide member governments with a powerful instrument to judge the effectiveness and international competitiveness of their higher education institutions, systems and policies' (Lederman 2007).

At first blush, there appear to be enormous methodological challenges with the project and indeed the results from the feasibility study are mixed. A technical advisory group has concluded that while in its opinion it is scientifically feasible to assess discipline-specific skills, there is less certainty around the reliability of the generic skills strand. In fact, the advisory group found that the questions used and based on the CLA 'proved excessively 'American' in an international context,' (OECD 2013, p. 24). This bears out criticisms that in trying to develop a common measurement tool that will compare diverse institutions with different missions and student populations, across a wide variety of countries, cultures and languages, the AHELO project imposes certain knowledge structures upon other cultures and histories. As Shahjahan (2013) has argued, AHELO is underpinned by particular, predominately Euro-American, values and knowledge systems that, in an echo of Gould's critique of biological determinism, are universalized and normalized in ways that mask over systemic inequities.

While the OECD insists AHELO will not be a ranking, it is difficult to see how it will be anything but, particularly when it is explicitly intended to help

governments benchmark the performance of their institutions against those in other jurisdictions. Once a measure is developed to assign a number to the performance of an institution or a program, whether based on research or teaching or some other category, the inevitable tendency will be for governments, institutions, and the media to rank results in a simplistic league table and to use those tables improperly. This is in fact precisely what happens now with the OECD's Programme for International Student Assessment (PISA), an international test given to 15 yr olds (Mortimore 2009). As is all too common in the sports world, a poor showing in the PISA standings, and ultimately perhaps in the AHELO league table, will unleash the usual barrage of external criticism and internal handwringing, accompanied by calls for sacking the coach and shaking up the team. No matter what the initial intention may be, the seemingly inevitable outcome of any measurement exercise in higher education is that the final results are used not to improve and support institutions and the people who work in universities and colleges by ensuring a supportive learning environment and respecting the professional autonomy of academic staff. Rather, it is to exert more external control.

Rankings as performance indicators

Rankings are close cousins to a host of key performance indicators, such as graduation rates, employment outcomes of graduates, faculty productivity measurements and citation performance, that have sprung up in recent decades (Bruneau & Savage 2002). Numerical evaluations and rankings of universities have been used by governments around the world for decades as tools to demand more for less, and to exert more control over university decision making. At a time when, in large parts of the world, public funding for higher education is stagnating or declining, students and their families are being asked to contribute more in the form of private fees, and the number of tenure and tenure-track faculty continues to decline rapidly in favor of fixed-term and precarious instructors (Altbach 2004), governments are attempting to use simplistic performance measures as justification for reduced resources and for more say over how institutions should spend their dwindling funding.

In this sense, university rankings can be dangerous. The high standards of quality we demand of higher education institutions may be undermined by

rankings because they can have the effect of compromising the quality of education by distorting institutional priorities. Some institutions, for instance, have simply manipulated data to improve their overall score (see Weisbrod & Asch 2009). Other universities and colleges, in an effort to boost their standing in league tables have focused internal resources on improving their performance on the evaluation criteria established by external rankings bodies (Marginson 2007). For instance, some Australian universities have employed full-time managers to work exclusively with ranking agencies and to develop strategies aimed at improving their position in the league tables (Trounson 2013). In other cases, institutions have adopted more selective admissions criteria for students. That may boost their reputational scoring, but it does so at the expense of providing greater equality of access (Clarke 2007).

In other instances, the quest to improve their ranking has meant that institutions have shifted internal resources towards more research intensive activities. Hazelkorn (2009b) notes that rankings have provoked some institutions to separate undergraduate and postgraduate teaching through the creation of semi-autonomous research institutes and graduate schools. The goal is to increase research intensity in order to produce more quantifiable outputs measured in the rankings. Consequently, rankings are influencing internal priorities by encouraging institutions to shift resources toward research-intensive disciplines such as applied science and engineering and away from the arts, humanities and social sciences which are 'deemed less vital to [an institution's] profile or perform poorly on comparative indicators,' (Hazelkorn 2009b, p. 62). The institutional obsession with rankings in turn affects internal decisions about the recruitment, promotion, and evaluation of staff with rankings now being used by some institutions to identify the 'best' and the 'under-performers' (Hazelkorn 2009b). The key point is that the underlying motivation behind the use of rankings in this way is not primarily to improve quality, but rather to elevate one's position in an arbitrary league table.

A further danger is that rankings can trump or even take the place of internal quality assurance practices. Given the diversity of institutions across the globe, a one-size-fits-all approach to assessing and measuring quality makes no sense. Rather than relying upon standardized and arbitrary ranking criteria to measure their performance, universities and the communities they serve would do better by developing their own quality assessment policies based upon their specific mission, size and budget. Rankings cannot

and should not be substitutes for robust quality assurance procedures. These procedures should encompass a rigorous assessment and review of the research, teaching and service at the program, departmental and institutional level. Reviews should be undertaken by academic peers and the focus should be on ensuring quality assurance is relevant to improving the full range of scholarly activities undertaken in higher education. Universities and their faculty have fought hard throughout history against attacks on their autonomy and independence. They have done so in order to preserve themselves as a unique place in society where, in the pursuit of knowledge, we can ask deeply disturbing questions and raise provocative challenges to existing beliefs. Handing over quality assessment to an external measurement system developed by rankings bodies is tantamount to giving up on one important fight in that long and continuing battle.

Numbers and rankings can move all too quickly and far too easily from relatively innocuous newspaper headlines and magazine covers into routine and instrumental assessments of institutions, departments, and even individual professors. Measurement criteria established by rankings, such as publication citations of faculty members, are now commonly used as simplistic performance indicators by governments and funding bodies to assess the relative effectiveness and competitiveness of their higher education institutions. In 2004, for example, the German government unveiled its Excellence Initiative, a controversial program that arose out of concerns that the country's universities were placing poorly in international rankings. The program assessed the research performance of universities based on the ranking criteria and rewarded those that came out on top of the league tables with additional funding (Labi 2010).

Rankings represent another symptom in the emergence of what has been identified as a new 'managerialism' in the academy (Deem 2008). The obsession with league tables, outcomes, targets, markets, benchmarks, and performance indicators have enormous consequences for the nature of academic work. When a university's standing is judged by the numerical research output of its faculty, the creativity of original research and discovery may be seen as secondary to the goal of improving a position in a league table. Similarly, an instructor's attempt to challenge the intellectual horizons of students may be subsumed beneath the view that teaching should provide 'consumers' of higher education — students and employers — only with what the market demands, employable skills.

The marketization of higher education

Given the myriad shortcomings and dangers associated with rankings, how do we explain their enduring popularity? Commercial rankings, with all their limitations and failings, are clearly in demand given the extent to which they help pad the profit margins of newspaper and magazine publishers. The obvious answer explaining the public appetite for rankings might be that the flourishing of rankings at both the national and international level reflects the way they are filling an information void that most universities and colleges have been unable to do. Would-be students considering their higher education options have few places to turn other than rankings, no matter how simple and superficial the results may be. Even so, there is surprisingly little evidence showing what effect rankings have on student choices. What most sociological research does reveal is that students are far more likely to base their higher education decisions on factors such as funding, proximity to an institution, and particular program offerings—just as was the case before rankings appeared (Gibbons & Vignoles 2009).

Part of a deeper answer to why rankings are so popular may lie in how their emergence has paralleled the rise of an increasingly consumerist and market-driven orientation within higher education globally. As Marginson & van der Wende (2007, p. 308) note: 'Global university rankings have cemented the notion of a world university competition or market capable of being arranged in a single 'league table' for comparative purposes and given a powerful impetus to intra-national and international competitive pressures in the sector.' Increasingly, higher education is a big business. Institutions, actively encouraged in many cases by government policy, openly and aggressively compete for a share of what is seen as a growing international 'market' in students. As Fig. 1 illustrates, the number of students studying abroad doubled from about 2 million in 2000 to over 4 million in 2010.

This market is made all the more attractive as governments cut university grants and allow tuition fees for international students to creep higher and higher. In nearly half of the OECD countries, international students pay far higher fees than their domestic counterparts. In Austria, for example,

the average tuition fees charged by public institutions for students who are not citizens of European Union or European Economic Area (EEA) countries are twice as high as for citizens of these countries. Similar policies are in place in Australia, Canada, Denmark (as of 2006 to 2007), Ireland, the Netherlands, New Zealand (except for doctoral students), Poland (only for public institutions), the Slovak Republic, Slovenia, Switzerland, Sweden (as of 2011), the United Kingdom and the United States of America (OECD 2011, p. 259). In this context, rankings feed this marketplace for international students, playing the role of consumer guide and marketing agent. Rankings provide seemingly independent information about the 'quality' of the product on offer from different suppliers. For their part, many institutions are increasingly obsessed with rankings not because they say anything meaningful about the quality of education, but primarily because they can help or hinder their attempts to increase market share.

This is not at all to say that we should not be collecting information or publicizing anything about universities and what they do. On the contrary, higher education institutions should be open and accountable. Normative statistics clearly have an important place in universities, helping guide policy makers, administrators and faculty to provide education more equitably and accessibly. Certain measurements are indispensable and may be used to show that universities need improved student financial assistance, more full-time faculty, and better equipment and facilities. But good universities and colleges can only ever in part be portrayed by indicators of various kinds.

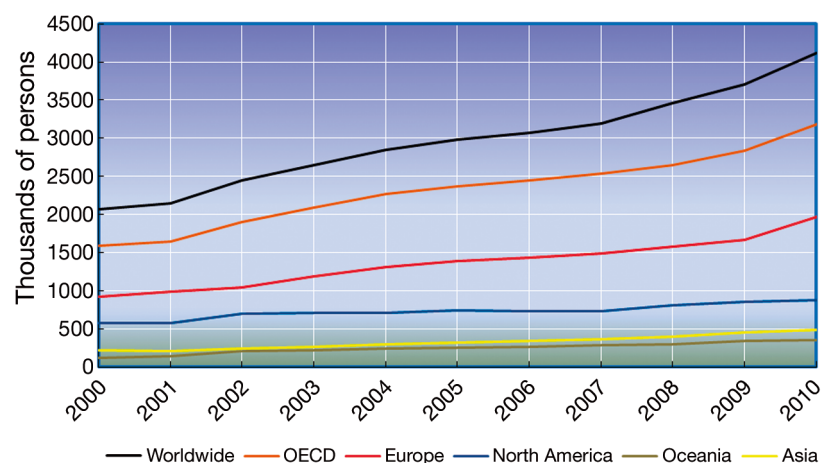


Fig. 1. Number of students enrolled outside their country of citizenship, by region of destination (2000 to 2010). Source: OECD Education at a Glance 2012 (www.oecd.org/edu/eag2012)

Conclusion

The academic community needs to reflect far more seriously and rigorously about what should count, how it should be counted, and to what end. There are, for instance, many things that faculty would consider essential to the provision of high quality education and research, such as: the commitment of institutions to academic freedom and free inquiry; a system of collegial governance in which faculty have a say over the educational decisions of the institution; internal quality assurance processes based upon peer review; access to professional development and support; and decent terms and conditions of employment. These are all essential elements of a supportive teaching and learning environment that promote quality, but are largely missing from the priorities of existing rankings.

Nevertheless, in criticizing rankings we must not limit ourselves to exposing their worst excesses. We should also guard against being satisfied with simply dreaming up new lists of statistics or refining existing measurements. There is a need to go much further in our thinking, well beyond the specific debate about rankings, to ask some basic but surprisingly elusive questions. What makes a 'good' university? What is it that we as societies ask of universities? What are the things about higher education that really matter? What is it that is of most value in higher education?

In her 2007 inauguration address as President of Harvard, Drew Faust provided us with an eloquent hint to some of the possible answers to these questions. Universities, she reminded her audience, are not academic factories whose products can be measured and tailored to market needs alone:

A university is not about results in the next quarter; it is not even about who a student has become by graduation. It is about learning that molds a lifetime, learning that transmits the heritage of millennia; learning that shapes the future. A university looks both backwards and forwards in ways that must—that even ought to—conflict with a public's immediate concerns or demands. Universities make commitments to the timeless, and these investments have yields we cannot predict and often cannot measure. [...] We are uncomfortable with efforts to justify these endeavors by defining them as instrumental, as measurably useful to particular contemporary needs. Instead we pursue them in part 'for their own sake,' because they define what has over centuries made us human, not because they can enhance our global competitiveness. (Faust 2007)

The proliferation of rankings and the debate they have generated highlight that higher education is today very much a contested terrain. Rankings stem from and encourage the belief of some that markets,

driven by the informed choices of consumers (formerly known as students) and by employers' demands for specific labor skills, should determine the curriculum, the teaching, and the research of universities. Rankings reinforce in not so subtle ways the view that higher education should be subordinated to short-term economic needs. In this way, what is at stake in the debate over rankings is not primarily a methodological dispute over whether this or that institution can be shown quantitatively to perform better or worse than others, or whether this or that ranking accurately measures quality. What really matters is whether higher education is to be thought of as having, as Drew Faust suggests, intrinsic value, or whether it is defined narrowly in terms of the crude instrumental logic of economic determinism and 'customer satisfaction'.

Acknowledgements. The author acknowledges the helpful assistance of Konstantinos Stergiou for the comments and direction he provided on earlier versions of this paper.

LITERATURE CITED

- Altbach P (2004) The deteriorating guru: the crisis of the professoriate. *Int High Educ* 36:2–3. Available at https://htmlbprod.bc.edu/prd/f?p=2290:4:0::NO:RP,4:P0_CONTENT_ID:100498
- Altbach P (2006a) The dilemmas of ranking. *Int High Educ* 42:2–3. Available at https://htmlbprod.bc.edu/prd/f?p=2290:4:0::NO:RP,4:P0_CONTENT_ID:100222
- Altbach P (2006b) The tyranny of citations. *Int High Educ* 43:3–5. Available at https://htmlbprod.bc.edu/prd/f?p=2290:4:0::NO:RP,4:P0_CONTENT_ID:100224
- Altbach P (2010) The state of rankings. *Inside Higher Ed*, 11 November 2010. Available at www.insidehighered.com/views/2010/11/11/altbach
- Birnbaum R (2001) *Management fads in higher education: where they come from, what they do, why they fail.* Jossey-Bass, San Francisco, CA
- Bruneau W, Savage DC (2002) *Counting out the scholars: the case against performance indicators in higher education.* James Lorimer, Toronto
- Clarke M (2007) The impact of higher education ranking on student access, choice, and ranking. *High Educ Eur* 32: 59–70
- Deem R (2008) Unraveling the fabric of academe: the managerialist university and its implications for the integrity of academic work. In: Turk JL (ed) *Universities at risk: how politics, special interests and corporatization threaten academic integrity.* James Lorimer, Toronto, p 256–281
- Faust D (2007) Installation address: unleashing our most ambitious imaginings. www.harvard.edu/president/installation-address-unleashing-our-most-ambitious-imaginings (accessed 8 March 2013)
- Gibbons S, Vignoles A (2009) *Access, choice and participation in higher education.* Centre for the Economics of Education, London School of Economics, London

- Gould SJ (1981) *The mismeasure of man*. W.W. Norton, New York, NY
- Guarino C, Ridgeway G, Chun M, Buddin R (2005) Latent variable analysis: a new approach to university ranking. *High Educ Eur* 30:147–165
- Hazelkorn E (2009a) Do university rankings measure what counts? *The Irish Times*, 21 April 2009. Available at <http://arrow.dit.ie/cgi/viewcontent.cgi?article=1022&context=cserart>
- Hazelkorn E (2009b) Rankings and the battle for world-class excellence: institutional strategies and policy choices. *High Educ Manag Policy* 21:55–76
- Labi A (2010) Germany pursues excellence over egalitarianism. *Chron High Educ* 27 June 2010. Available at <http://chronicle.com/article/Germany-Pursues-Excellence/66048/>
- Lederman D (2007) A worldwide test for higher education? *Inside Higher Ed* 19 September 2007. Available at www.insidehighered.com/news/2007/09/19/international
- Marginson S (2007) Australia: the power of rankings. *University World News*, 11 November 2007, Issue No. 5. Available at www.universityworldnews.com/article.php?story=20071108143650369
- Marginson S, van der Wende M (2007) To rank or to be ranked: the impact of global rankings in higher education. *J Stud Int Educ* 11:306–329
- Mortimore P (2009) Alternative models for analysing and interpreting countries' performance in PISA. Education International Research Institute, Brussels
- OECD (Organization for Economic Cooperation and Development) (2010) Group of national experts on the AHELO feasibility study: AHELO assessment design. OECD, Paris. Available at [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne\(2010\)17&doclanguage=en](http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2010)17&doclanguage=en)
- OECD (2011) *Education at a glance 2011: OECD indicators*. OECD, Paris
- OECD (2013) *Assessment of higher education learning outcomes feasibility study report, Vol 2. Data analysis and national experiences: executive summary*. OECD, Paris. Available at www.oecd.org/edu/skills-beyond-school/AHELO%20FS%20Report%20-%20Volume%20%20Executive%20Summary.pdf
- Shahjahan RA (2013) Coloniality and a global testing regime in higher education: unpacking the OECD's AHELO initiative. *J Educ Policy*, doi:10.1080/02680939.2012.758831
- Tremblay K, Lalancette D, Roseveare D (2012) *Assessment of higher education learning outcomes: feasibility study report, Vol 1. Design and implementation*. OECD, Paris. Available at www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf
- Trounson A (2013) Universities hire rankings pros. *Inside Higher Ed*, 20 March 2013. Available at www.insidehighered.com/news/2013/03/20/australian-universities-dedicate-positions-working-rankings-groups
- Weisbrod BA, Asch ED (2009) Gaming the rankings. *Inside Higher Ed*, 9 June 2009. Available at www.insidehighered.com/views/2009/06/09/asch

Editorial responsibility: Konstantinos Stergiou, Thessaloniki, Greece

*Submitted: April 2, 2013; Accepted: May 14, 2013
Proofs received from author(s): June 19, 2013*