Inconsistent year-to-year fluctuations limit the conclusiveness of global higher education rankings for university management

Johannes Sorz, Bernard Wallner, Horst Seidler, Martin Fieder

Backround University rankings are getting very high international media attention, this holds particularly true for the Times Higher Education Ranking (THE) and the Shanghai Jiao Tong University's Academic Ranking of World Universities Ranking (ARWU). We therefore aimed to investigate how reliable the rankings are, especially for universities with lower ranking positions, that often show inconclusive year-to-year fluctuations in their rank, and if these rankings are thus a suitable basis for management purposes. Methods We used the public available data from the web pages of the THE and the ARWU ranking to analyze the dynamics of change in score and ranking position from year to year and we investigated possible causes for inconsistent fluctuations in the rankings by the means of regression analyses. Results Regression analyses of results from the THE and ARWU from 2010-2014 show inconsistent fluctuations in the rank and score for universities with lower rank positions (below position 50) which lead to inconsistent "up and downs" in the total results, especially in the THE and to a lesser extent also in the ARWU. In both rankings the mean year-to-year fluctuation of universities in groups of 50 universities aggregated by descending rank increases from less than 10% in the group of the 50 highest ranked universities to up to 60% in the group of the lowest ranked universities. Furthermore year-to-year results do not correspond in THES- and ARWU-Rankings for universities below rank 50. Discussion We conclude that the observed fluctuations in the THE do not correspond to actual university performance and ranking results are thus of limited conclusiveness for the university management of universities below a rank of 50. While the ARWU rankings seems more robust against inconsistent fluctuations, its year to year changes in the scores are very small, so essential changes from year to year could not be expected. Furthermore year-to-year results do not correspond in THES- and ARWU-Rankings for universities below rank 50. Neither the THES nor the ARWU offer great value as a tool for university management in their current forms for universities ranked below 50, thus we suggest that both rankings alter their ranking procedure insofar as universities below position 50 should be ranked summarized only in groups of 25 or 50. Additionally the THE should omit the peer reputation survey, which most likely contributes heavily to the inconsistent year-to-

year fluctuations in ranks, and ARWU should be published less often to increase its validity.

1   Inconsistent year-to-year fluctuations limit the conclusiveness of global

2   higher education rankings for university management

5   Johannes Sorz[1], Bernard Wallner[2,3], Horst Seidler[2] and Martin Fieder[2]

7   [1]University of Vienna, Office of the Rectorate, Universitätsring 1, A-1010 Vienna (Austria)

8   [2]University of Vienna, Department for Anthropology, Althanstrasse 14, A-1090 Vienna (Austria)

9   3 University of Vienna, Unit for Quality Assurance, Universitätsstraße 5, 1010 Vienna (Austria)

11   **Backround**

12   University rankings are getting very high international media attention, this holds particularly

13   true for the Times Higher Education Ranking (THE) and the Shanghai Jiao Tong University's

14   Academic Ranking of World Universities Ranking (ARWU). We therefore aimed to investigate

15   how reliable the rankings are, especially for universities with lower ranking positions, that often

16   show inconclusive year-to-year fluctuations in their rank, and if these rankings are thus a suitable

17   basis for management purposes.

18   **Methods**

19   We used the public available data from the web pages of the THE and the ARWU ranking to

20   analyze the dynamics of change in score and ranking position from year to year and we

21   investigated possible causes for inconsistent fluctuations in the rankings by the means of

22   regression analyses.

23   **Results**

24   Regression analyses of results from the THE and ARWU from 2010-2014 show inconsistent

25   fluctuations in the rank and score for universities with lower rank positions (below position 50)

26 which lead to inconsistent "up and downs" in the total results, especially in the THE and to a

27 lesser extent also in the ARWU. In both rankings the mean year-to-year fluctuation of

28 universities in groups of 50 universities aggregated by descending rank increases from less than

29 10% in the group of the 50 highest ranked universities to up to 60% in the group of the lowest

30 ranked universities. Furthermore year-to-year results do not correspond in THES- and ARWU-

31 Rankings for universities below rank 50.

32

33 **Discussion**

34 We conclude that the observed fluctuations in the THE do not correspond to actual university

35 performance and ranking results are thus of limited conclusiveness for the university

36 management of universities below a rank of 50. While the ARWU rankings seems more robust

37 against inconsistent fluctuations, its year to year changes in the scores are very small, so essential

38 changes from year to year could not be expected. Furthermore year-to-year results do not

39 correspond in THES- and ARWU-Rankings for universities below rank 50. Neither the THES

40 nor the ARWU offer great value as a tool for university management in their current forms for

41 universities ranked below 50, thus we suggest that both rankings alter their ranking procedure

42 insofar as universities below position 50 should be ranked summarized only in groups of 25 or

43 50. Additionally the THE should omit the peer reputation survey, which most likely contributes

44 heavily to the inconsistent year-to-year fluctuations in ranks, and ARWU should be published

45 less often to increase its validity.

46

47 **Introduction**

48 Global higher education rankings have received much attention recently and, as can be witnessed

49 by the growing number of rankings being published every year, this attention is not likely to

50 subside. Besides the arguable use of ranking results as an instrument for university management,

51 it is still a common practice in many universities to use rankings as an indicator for academic

52 performance. Rankings became a big business and as of today a plethora of regional and national

53 rankings exist, advocated by their publishers as potentially efficient and effective means of

54   providing needed information to universities on areas needing improvement (Dill and Soo,

55   2005). Numerous studies have analyzed and criticized higher education rankings and their

56   methodologies (van Raan, 2005; Buela-Casal et al., 2007; Ioannides et al., 2007; Hazelkorn,

57   2007; Aguillo et al., 2010; Benito and Romera 2011; Hazelkorn 2011; Rauhvargers, 2011;

58   Tofallis, 2011; Saisana et al. 2011; Safon, 2013; Rauhvargers, 2013; Bougnol and Dulá, 2014).

59   This casts justified doubt on a sensible comparison of universities hailing from different higher

60   education systems and varying in size, mission and endowment based on mono-dimensional

61   rankings and league tables and hence on the usability of such rankings for university

62   management and policy making (O'Connell, 2013; Hazelkorn, 2014). Several studies have

63   demonstrated that data used to calculate ranking scores can be inconsistent. Bibliometric data

64   from international databases (Web of Science, Scopus), used in most global rankings to calculate

65   research output indicators, favor universities from English-speaking countries and institutions

66   with a narrow focus on highly-cited fields, which are well covered in THEe databases. This puts

67   universities from non-English-speaking countries, with a focus on the arts, humanities and social

68   sciences, at a disadvantage when being compared in global rankings (Calero-Medina et al., 2008;

69   van Raan et al., 2011; Waltman et al., 2012). Data submitted by universities to ranking agencies

70   (e.g. personnel data, student numbers) can be problematic to compare due to different standards.

71   These incompatibilities are being amplified because university managers have become

72   increasingly aware of global rankings and try to boost their performance by "tweaking" the data

73   they submit to the ranking agencies (Spiegel Online, 2014). Beyond all the data issues, there is

74   the effect that universities with lower ranking positions often encounter volatile ups and downs

75   in their consecutive year-to-year ranks. These effects make university rankings an inconclusive

76   tool for university managers: the ranking results simply do not reflect the universities' actual

77   performance or their management strategies. Ranking results need to be consistent to be of use,

78   so that long-term strategies (e.g. the hiring of high-calibre researchers from abroad or

79   investments in doctoral education) are reflected in year-to-year scores and ranks and in perennial

80   trends. Furthermore, results from various rankings should be concordant to allow a sort of meta-

81   analysis of rankings.

82

83   Bookstein et al. (2010) found unacceptably high year-to-year variances in the score of lower

84   ranked universities in the THE, Jovanovic et al. (2012) and Docampo (2013) found a large

85 number of fluctuations and inconsistencies in the ranks of the ARWU. As we again observed
86 puzzling results in the THE 2014-15 and the ARWU 2014, that were both published recently, we
87 accordingly, analyzed the fluctuations in score and rank of the THE and the ARWU. By
88 calculating regression analyses for both rankings for consecutive years for 2010-2014 we tried to
89 determine the amount of inconsistent fluctuations that can most likely not be explained by
90 changes in university performance (e.g. by increase of publications/citations, change in
91 student/faculty numbers) and listed the universities with the most extreme changes in ranking
92 position for the THE and ARWU. Furthermore, the mean percentages of universities that
93 changed their rank in their groups were calculated for groups of 50 universities aggregated by
94 descending rank in both THE and ARWU and we calculated a regression of the ranking positions
95 of the first 100 universities in the THE 2014 on the first 100 universities in the ARWU 2014

96

97 THE
98 The methodology of the THE was revised several times in varying scale, before and after the
99 split with Quacquarelli Symonds (QS) in 2010 and the new partnership with Thompson Reuters.
100 THE calculates 13 performance indicators, grouped into the five areas Teaching (30%), Research
101 (30%), Citations (30%), Industry income (2.5%) and International outlook (7.5%). However,
102 THE does not publish the scores of individual indicators, only those of all five areas combined.
103 Since 2010, the research output indicators are calculated based on Web of Science data. Most of
104 the weight in the overall score is made up by the normalized average citations per published
105 paper (30%), and by the results of an academic reputation survey (33%) assessing teaching and
106 research reputation and influencing the scores of both areas (Rauhvargers, 2013; THE, 2014). In
107 the past, criticism has been levied against this survey. Academic peers can choose universities in
108 their field from a preselected list of institutions and, although universities can be added to the
109 list, those present on the original list are more likely to be nominated. This leads to a distribution
110 skewed in favor of the institutions at the top of the rankings (Rauhvargers, 2011 and 2013). THE
111 allegedly addressed this issue by adding an exponential component to increase differentiation
112 between institutions, yet no information is available on its mode of calculation (Baty, 2011 and
113 2012).

114

115 ARWU

116 ARWU ranks more than 1000 (of ca. 17,000 universities in the world) and publishes the best 500

117 on the web. In addition ARWU offers to field rankings that cover several subjects and subject

118 rankings for Mathematics, Physics, Chemistry, Computer Science and Economics & Business.

119 Universities are ranked according to their research performance, including alumni (10%) and

120 staff (20%) winning Nobel Prizes and Fields Medals, highly cited researchers in in 21 broad

121 subject categories in the Web of Science (20%), papers published in Nature and Science (20%),

122 papers indexed in major citation indices (20%), and the per capita academic performance of an

123 institution (10%). Calculation of indicators remained relatively constant since 2004. ARWU

124 ranks universities individually or into bands by sorting on the total score which is the linearly

125 weighted sum of the six research output indicator scores derived from the corresponding raw

126 data by transformations. Institutional data (number of academic staff) is not provided by

127 universities but obtained from national agencies such as ministries, national bureaus and

128 university associations (ARWU, 2013). In contrast to the THE, there are no teaching/student

129 related indicators or any peer survey component in the ARWU. Due to reliance on ISI subject

130 fields the areas of natural sciences, medicine and engineering dominate the citation indicator,

131 putting universities with a focus on the arts, humanities and social sciences. The per capita

132 performance is the only ARWU indicator that takes into account the size of the institution, thus

133 small but excellent institutions have less of a chance to perform well in the ARWU-Ranking

134 (Rauhvargers, 2011). Already several studies, i.e. Docampo (2011 and 2013) analyzed the

135 ARWU and its indicators and found inconsistencies and unwanted dynamical effects. We have

136 no further information beside the public available information on the indicators and their weights

137 on how the scores are calculated for THE and ARWU.

138

139 **Methods**

140 We used the publicly available data on scores and ranks from the THE and ARWU for the years

141 2010, 2011, 2012, 2013 and 2014, including in the THE all universities ranked between 1 and

142 200 and in the ARWU the universities ranked between 1 und 100, as ARWU starts aggregating

143 the ranking from rank 101 on. We performed the following analysis for both rankings: i) we

144 plotted and regressed the scores of the rankings of the year t-1 on the scores of the year t; ii) we

145 plotted and regressed the ranks of the rankings of the year t-1 on the ranks of the year t; iii) we

146 plotted the associations of scores and ranks and approximated the function of the association

147   between scores and ranks iv), we investigated the concordance (ranking position of the first 100

148   universities) of the THE ranking with the ARWU ranking. For this purpose we regressed the

149   position of the first 100 universities in the THE-Ranking (2014-15) on the ranking position of the

150   first 100 university in the ARWU-Ranking (2014) and finally vii) to include also universities

151   ranked below 200 we aggregated the THE from "position 1" on in steps of 50 universities, i.e.

152   we defined 8 aggregated ranking groups (1-50, 51-100, 101- 150, 151- 200, 201- 250, 251- 300,

153   301- 350, 351- 400) for the THE. As the ARWU starts aggregating in steps of 100 universities

154   from rank 201 on, we refrained from including universities ranked lower than 200 in the ARWU

155   in our analysis, due to comparability; so therefore for the ARWU we defined 4 aggregated

156   ranking groups of also 50 universities (1-50, 51-100, 101- 150, 151- 200).

157   On basis of this rearrangement we made the following calculations: we calculated the year-to-

158   year mean fluctuations (%), thus the percentage of universities that changed their rank beyond

159   their aggregated ranking groups (moving upwards, moving downwards, being newly in the rank,

160   respectively dropping out of the ranking) to get an estimate of yearly fluctuation according to the

161   ranking groups. We calculated the mean change of years 2012, 2013 and 2014. Direction and

162   amount of change were not considered, only the fact that a university did change ranking group,

163   i. e, moving within the ranking, dropping out of the ranking, respectively coming newly into the

164   ranking, was counted.

165

166

167

168

169

170   **Results**

171

172   THE:

173   Regression of the scores and ranks of two consecutive years.

174   The regression of the scores – particularly of the ranking 2010-2011 regressing on the scores of

175   the ranking of 2011-2012 – shows very high fluctuations (figure 1a), especially for the lower

176   ranked universities. Moreover, the fluctuations among the lower ranked universities seem to be

177   higher compared to the THE performed by QS before 2010 (Bookstein et al. 2010, figure 1).

178   Note that in the rankings in the years following 2010-2011, the fluctuations in the THE-Ranking

179   did decrease (figure 1b-d). Supplementary tables 1 a- h) show the regression models including

180   regression coefficients, confidence intervals, p- values, degrees of freedom and R^2 values for

181   each model. R^2 ranges from 0.71 to 0.98, i.e. the models with the small R^2 do explain less

182   variance. This confirms the visual impression of particularly high year-to-year fluctuations as

183   depicted in figures 1a) and 1e). Also when universities are aggregated in groups of 50

184   universities by decreasing rank, the observed fluctuations increase with increasing rank from

185   year to year and the mean year-to-year fluctuation (mean percentages of universities that

186   changed their rank in their group) increases in each subsequent lower ranked group. (figure 3).

187   While the mean year-to-year fluctuation is less than 10% in the first group (1-50) it is over 60%

188   in the lowest ranked universities group (351- 400). The most extreme cases of "fluctuating

189   university ranks" from 2010 to 2014 in the THE rankings are displayed in table S3.

190

191

192   Association between Scores and Ranks

193

194   A general problem of the THE remains: the difference in the scores among the 50 highest scoring

195   universities is considerably higher compared to the difference among the lower scoring

196   universities. This clearly suggests a non-linear relationship between scores and ranks (figure 2 a-

197   e). The consequence is that the ranks of the high scoring universities are much more robust to

198   deviations in the scores from year to year. In the lower ranking universities, however, even very

199   small, more or less random deviations (around 0.5%) lead to unexpected "high jumps" in the

200   ranks from year to year (figure 1e-h). We assume, that these fluctuations are to a large extent

201   caused by the results of the peer reputation survey, which can be skewed by low response rates

202   (Rauhvargers, 2011) and the "Matthew effect" (Merton, 1968). Interestingly, the association

203   between scores and ranks approximates pretty well a power function of the form rank = score +

204   score[b], there b ranges between -4.106 and -4.88 for the THE ranking. In supplementary figure 1),

205   the fit on basis of a power function for the ranking 2014 is shown exemplarily (as the figures are

206   quite similar we did refrain from plotting all the power fits).

207

208

209   ARWU:

210   While still on a high level, the regression of ranks and scores of the ARWU, show much less

211   fluctuations compared with the THE. This indicates a more robust set of indicators. Furthermore

212   the ARWU shows a similar, but even a more extreme pattern of non-linearity between ranks and

213   scores, compared with the THE. Particularly the first ranked university, Harvard University,

214   scores far ahead of all the other universities in the ARWU at each year. As in the THE the

215   association between ranks and scores flattens from rank of 50 on (figure 4a-e). As in the THE the

216   non-linear relationship between ranks and scores increases the fluctuations in the ranking

217   positions of the universities ranked approx. below 50 from year to year (figure 3a-d).

218   Supplementary tables 2 a- h) show the regression models including regression coefficients,

219   confidence intervals, p- values, degrees of freedom and R^2 values for each model. R^2 ranges from

220   0.71 to 0.98, i.e. the models with the small R^2 do explain less variance, the amount of

221   inconsistent fluctuations seem to be higher. The higher R^2 values indicate, that the ARWU shows

222   less inconsistent fluctuations compared to the THE, which confirms the visual impression of the

223   figures 3 a- h) and confirms our notion that the inconsistent fluctuations in the THE are largely

224   caused by the results of the peer reputations survey, which is not included in the ARWU. As in

225   the THE the associations scores and ranks approximates pretty well a power function of the form

226   rank = score + score^b, there b ranges between -2.96 and -3.00 for the ARWU (compared with -

227   4.106 and -4.88 for the THE).

228   As in the THE, if universities are aggregated in steps of 50 universities by decreasing rank, the

229   fluctuations increase with increasing rank from year to year and the mean year-to-year

230   fluctuation (%) increases in lower ranking groups (figure 6). While the mean year-to-year

231   fluctuation is less than 10% in the first group (1-50) it is over 40% in the lowest ranked

232   universities group (151-200). The most extreme cases of "fluctuating university ranks" from

233   2010 to 2014 in the ARWU rankings, are displayed in table S4.

234   Correlation between THE and ARWU

235

236   A really dramatic amount of inconsistent fluctuations reveal the regression of the ranks in the

237   THE on the ranks in the ARWU: for the universities ranked approximately lower the 50th rank

238   there is virtually no correlation between the THE and the ARWU (figure 7). Regression could

239   only be plotted for universities ranked in both rankings among the first 100). The R^2 of 0.52

240   indicates, as seen in the plot (figure 7), that only a relatively small amount of the variance can be

241   explained by the association between the THE and the ARWU (supplement table S3). i. e. the

242   inconsistent fluctuations are quite high. This implies, that the THE and ARWU make substantial

243   different statements on the performance of universities ranked below 50, thus making both

244   rankings hard if not impossible to compare. While this effect is to some degree understandable,

245   due to the different setup of indicators in both rankings, however one could assume that the

246   reflection of academic performance of universities is reflected more homogenously in both

247   rankings. More homogenous results would also allow a "meta-ranking" which could be of more

248   value for university management than singular contradicting ranking results.

249

250

251

252

253

Discussion

254

255 High ranking positions achieved by a small group of universities are often self-perpetuating,

256 especially due to the intensive use of peer review indicators, which improve chances of

257 maintaining a high position for universities already near the top (Bowman and Bastedo, 2011;

258 Rauhvargers, 2011). This phenomenon also corresponds to the "Matthew effect", which was

259 coined by Merton (1968) to describe how eminent scientists will often get more credit than a

260 comparatively unknown researcher, even if their work is similar: credit will usually be given to

261 researchers who are already famous. The intensive and exaggerated discussion in the media of

262 the "up and downs" of universities in the THE is particularly misleading for universities with

263 lower ranking positions (below approximately a score of 65% and a rank of 50; above scores of

264 65%, the relationship between ranks and scores is steeper, and it flattens for scores below 65%).

265 This is because the ranking positions suggest substantial shifts in university performance despite

266 only very subtle changes in score. In fact, merely random deviations must be assumed. One

267 reason lies in the weighing of indicators by THE, with the emphasis on citations (30% of the

268 total score) and the peer reputation survey (33% of the total score). For lower ranked

269 universities, a few highly cited publications, or the lack thereof, or few points asserted by peers

270 in the reputation survey, probably make a significant difference in total score and position.

271

272 Ranking results have a major influence on the public image of universities and can even impact

273 their claim to resources (Espeland & Saunder, 2007; Hazelkorn, 2011). Accordingly, inconsistent

274 fluctuations in ranking positions can have serious implications for universities, especially when

275 the media or stakeholders interpret them as direct results of more or less successful university

276 management. The use of monodimensional rankings for university management is generally

277 doubtful. Our results show, that especially the THE in its current form has very limited value for

278 the management of universities ranked below 50. This is because the described fluctuations in

279 rank and score probably do not reflect actual performance, whereby the results cannot be used to

280 assess the impact of long-term strategies. "Rankings are here to stay, and it is therefore worth the

281 time and effort to get them right" warns Gilbert (2007).What could be done to address the

282 fluctuations in the THE for universities below rank 50 to make it a more usable tool to assess

283 actual performance for university management? THE has already addressed fluctuations to some

284 extent by ranking universities only down to position 200, followed by groups of 25 from 201-300

285 and groups of 50 from 300 to 400. Nonetheless, based on our data we believe that this is not

286 going far enough and suggest that universities should be summarized in groups of 25 or 50 below

287 the position of 50. Furthermore, we believe that these inconsistent fluctuations are caused to a

288 large extent by the results of the peer reputation survey, which can be skewed by low response

289 rates (Rauhvargers, 2011) and biased by the "Matthew effect" that favours already well-

290 renowned institutions (Merton, 1968). The latter could possibly also help to explain the

291 consistency in the scores and ranks in the group of the top-50 universities. Thus, in order to

292 increase the validity of the THE, the peer reputation survey should be omitted or given less

293 weight in future rankings.

294 The analysed curves of scores vs. ranking positions in figure 2 do have analogous characteristics

295 for example to semi-logarithmic curves produced in analytic biochemistry. The accuracy of such

296 curves is limited to the steepest slope of the curve, whereas asymptote areas deliver higher

297 fuzziness (Chan, D.W. (ed), 1992). Thus, a further suggestion to avoid the blurring dilemma is

298 the methodological approach of introducing a standardization process for THE data. This would

299 involve using common suitable reference data to create calibration curves represented by non-

300 linearity or linearity. Simple calibration would be as suggested to categorize universities in

301 ranking groups or to apply a transformation to the scores, such as a logarithmical transformation.

302 Comparing the year to year fluctuation in the ARWU with the THE reveals, that fluctuation in

303 the ARWU ranking is overall lower as in the THE ranking (figure 1 vs. figure 3), i.e. the ARWU

304 ranking seems to be more stable. This is on one hand a good message: a smaller amount of

305 fluctuations, but on the other hand, it has to be asked if a yearly publication of the ARWU makes

306 sense, if no "real" changes can be expected. However the same holds true for all rankings

307 published on a yearly basis: no factual changes reflecting university strategies can be expected.

308

309 The astonishing low correlation between the ranks of the THE and the ARWU ranking,

310 particularly for the universities ranked below 50 in both rankings, creates another serious doubt

311 if rankings should be used for any management purposes at all. Maybe a "meta- analysis" of

312 rankings could be reasonable to derivate consistent and reliable results from rankings. If done,

313 such and meta- analysis should include as many rankings as possible to reduce the amount of

314 inconsistent fluctuations that do not reflect actual university performance.

315

316 **Conclusion**

317 Both rankings show fluctuations in the rank and score particularly for universities with lower

318 ranking position (below position 50) which lead to inconsistent "up and downs" in the total

319 results, especially in the THE and to a lesser extend also in the ARWU. The observed

320 fluctuations do most likely not correspond to academic performance, so therefore neither the

321 THES nor the ARWU offer great value as a tool for university management in their current

322 forms for universities ranked below 50.

323
324

325

326

327 **References**

328

329 Aguillo, I.F., Bar-Ilan, J., Levene, M. and Ortega, L.J. (2010) Comparing university rankings.

330   Scientometrics 85, 243–256

331 ARWU (2013) The Academic Ranking of World Universities – Methodology

332   (http://www.shanghairanking.com/ARWU-Methodology-2013.html, Retrieved on 23.02.

333   2015)

334 Benito, M. and Romera, N. (2011) Improving quality assessment of composite indicators in

335   university rankings: a case study of French and German universities of excellence.

336   Scientometrics (2011) 89, 153–176

337 Baty, P. (2011) "THE Global Rankings: Change for the better", Times Higher Education, 6

338   October. Retrieved on 11 November 2014 from:

339   http://www.timeshighereducation.co.uk/world-universityrankings/2011-12/world

340   ranking/methodology

341 Baty, P. (2012), "The essential elements in our world-leading formula", Times Higher

342   Education, 4 October. Retrieved on 11. November 2014 from:

343   http://www.timeshighereducation.co.uk/worlduniversity-rankings/2012-13/world-

344   ranking/methodology

345   Billaut, J. C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shangai ranking: An

346   MCDM view. Scientometrics, 84(1), 237–263.

347   Bowman, A.M. and Bastedo, N.M. (2011) Anchoring effects in world university rankings:

348   exploring biases in reputation scores. Higher Education 61, 431–444

349   Bookstein, F.L, Seidler, H., Fieder, M., Winckler, G. (2010) Too much noise in the Times

350   Higher Education rankings. Scientometrics 85, 295–299

351   Buela-Casal, G., Gutierrez-Martinez, O., Bermudes-Sanchez, M., Vadillo-Munoz, O. (2007)

352   Comparative study of international academic rankings of universities. Scientometrics 71,

353   349-365

354   Bougnol, M.L., Dula, J.H. (2014) Technical pitfalls in university rankings, Higher Education,

355   DOI 10.1007/s10734-014-9809-y

356   Calero-Medina, C., López-Illescas, C., Visser, M.J and Moed, H.F. (2008) Important factors

357   when interpreting bibliometric rankings of world universities: an example from oncology.

358   Research Evaluation, 17(1), 71–81

359   Chan, D.W. (ed) (1992) Immunoassay automation: a practical guide. Academic Press: San Diego

360   Dill, D.D, Soo, M. (2005) Academic quality, league tables, and public policy: A cross-national

361   analysis of university ranking systems. Higher Education, 49, 495–533

362   Docampo, D. (2011). On using the Shanghai ranking to assess the research performance of

363   university systems. Scientometrics, 86(1), 77–92.

364   Docampo, D. (2013). Reproducibility of the Shanghai academic ranking of world universities

365   results. Scientometrics, 94(2), 515–524.

366   Espeland, W.N and Sauder, M. (2007) Rankings and Reactivity: How Public Measures Recreate

367   Social Worlds, American Journal of Sociology 113 (1), 1–40

368   Gilbert, A. (2007). Academics strike back at spurious rankings. Nature 447, 514–515

369   Hazelkorn, E., (2007) Impact and influence of league tables and ranking systems on higher

370   education decision-making. Higher Education Management and Policy 19 (2), 87–110

371   Hazelkorn, E. (2011) Rankings and the Reshaping of Higher Education: the battle for world class

372   excellence (Basingstoke, Palgrave-MacMillan).

373   Hazelkorn, E. (2014) Reflections on a Decade of Global Rankings: what we've learned and

374   outstanding issues, European Journal of Education, 49 (1), 12–28

375 Ioannidis, J. P. A., Patsopoulos, N. A., Kavvoura, F. K., Tatsioni, A., Evangelou, E., Kouri, I.,
376   Contopoulos-Ioannidis D.P., Liberopoulos G.. (2007) International ranking system for
377   universities and institutions: A critical appraisal. BMC Medicine, 5, 30.

378 Jovanovic, M., Jeremic, V., Savic, G., Bulajic, M., Martic, M. (2012) How does the
379   normalization of data affect the ARWU-ranking? Scientometrics (2012), 93,319–327

380 Liu, N. C., Cheng, Y. (2005). Academic ranking of world universities: Methodologies and
381   problems, Higher Education in Europe, 30(2), 127–136.

382 Merton, R.K. (1968) The Matthew effect in science. Science 159, 56-63

383 O'Connell, C. (2013) Research discourses surrounding global university rankings: exploring the
384   relationship with policy and practice recommendations, Higher Education, 65, 709–723

385 Rauhvargers, A. (2011) EUA Report on Global Rankings and their Impact – Report I (European
386   University Association)

387 Rauhvargers, A. (2013) EUA Report on Global Rankings and their Impact – Report II (European
388   University Association)

389 Safon, V. (2013) What do global university rankings really measure? The search for the X factor
390   and the X entity. Scientometrics 97, 223–244

391 Saisana, M., d'Hombres, B., Saltelli X. (2011) A Rickety numbers: Volatility of university
392   rankings and policy implications. Research Policy 40, 165–177

393 Spiegel Online (2014). Deutsche Unis im "THE"-Ranking: Das Wunder von Tübingen.
394   02.10.2014, Retrieved on 11. November 2014:
395   http://www.spiegel.de/unispiegel/studium/uni-ranking-hochschulenim- the-ranking-a-
396   994684.html

397 Times Higher Education (2014) World University Rankings 2014-2015 methodology
398   http://www.timeshighereducation.co.uk/world-university-rankings/2014-
399   15/worldranking/methodology

400 Tofallis, C. (2012) A different approach to university rankings. Higher Education 63, 1 –18

401 van Raan, T. (2005) Fatal attraction: Conceptual and methodological problems in the ranking of
402   universities by bibliometric methods. Scientometrics, 62(1), 133–143

403 van Raan, T., Leeuwen, T., and Visser, M. (2011) Severe language effect in university rankings:
404   particularly Germany and France are wronged in citation-based rankings. Scientometrics
405   88, 495–498

406   Waltman, L., Calero-Medina, C., Kosten, J., Noyons, N.C., Tijssen, R.J., and Wouters, P. (2012)

407   The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. CWTS

408   Working Paper Series. http://arxiv.org/abs/1202.3941

1

Figure 1

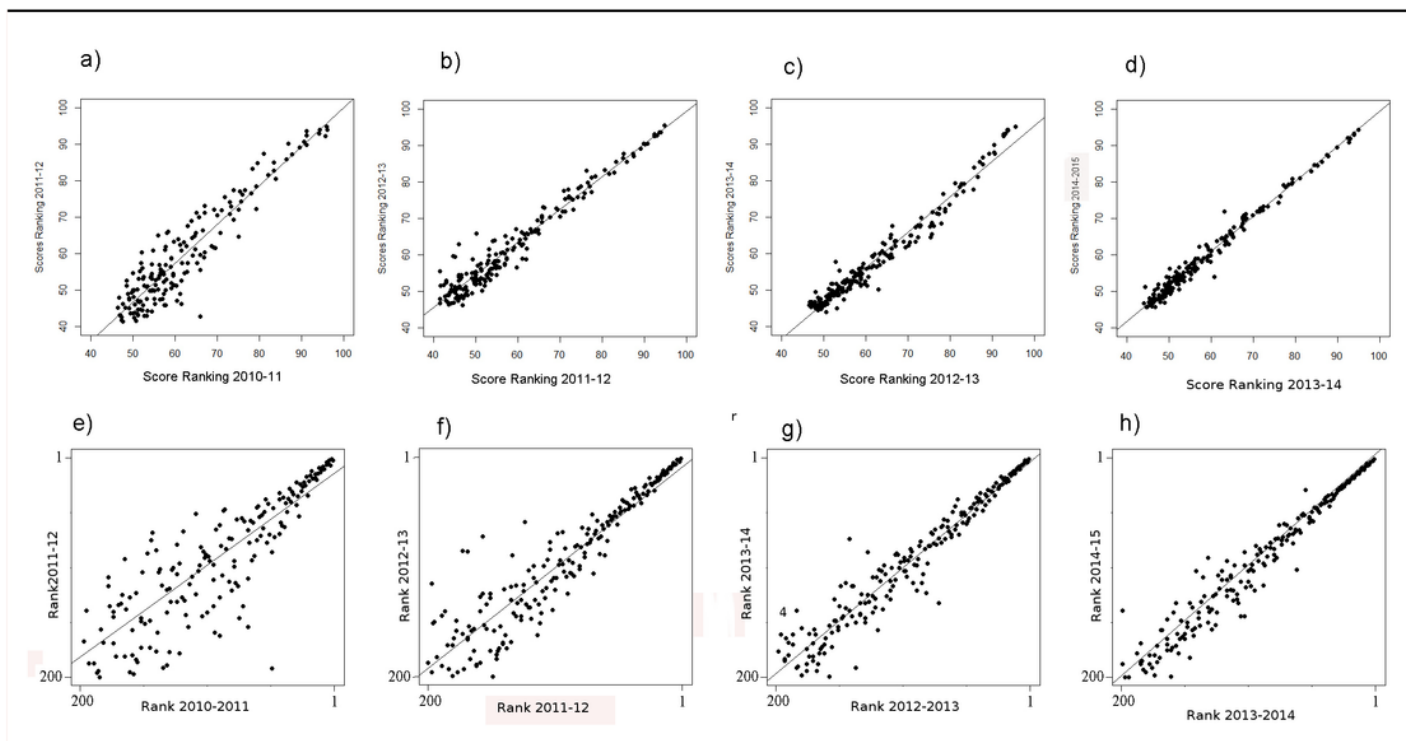
Figure 1a-1h) THE: Scores of the year t regressing on the score of the year t-1 from the ranking 2010-11 on. Figure 1a-1d). Ranks of the year t regressing on the ranks of the year t-1 from the ranking 2010-11 on (1e-1h). Linear regression line indicates perfect association, e.g. no changes in ranks and scores between two consecutive rankings. Remark: THE denotes the rankings in academic years (figures 1 & 2), whereas ARWU denotes the rankings in calendar years (figure 4 & 5).

2

Figure 2

Figure 2 a-e). Ranks plotted against scores for the THE a) score 2010- 2011 plotted against rank 2010- 2011; b) score 2011- 2012 plotted against rank 2011- 2012; c) score 2012- 2013 plotted against rank 2012- 2013; d) score 2013- 2014 plotted against rank 2013- 2014 and e) score 2014- 2015 plotted against rank 2014- 2015

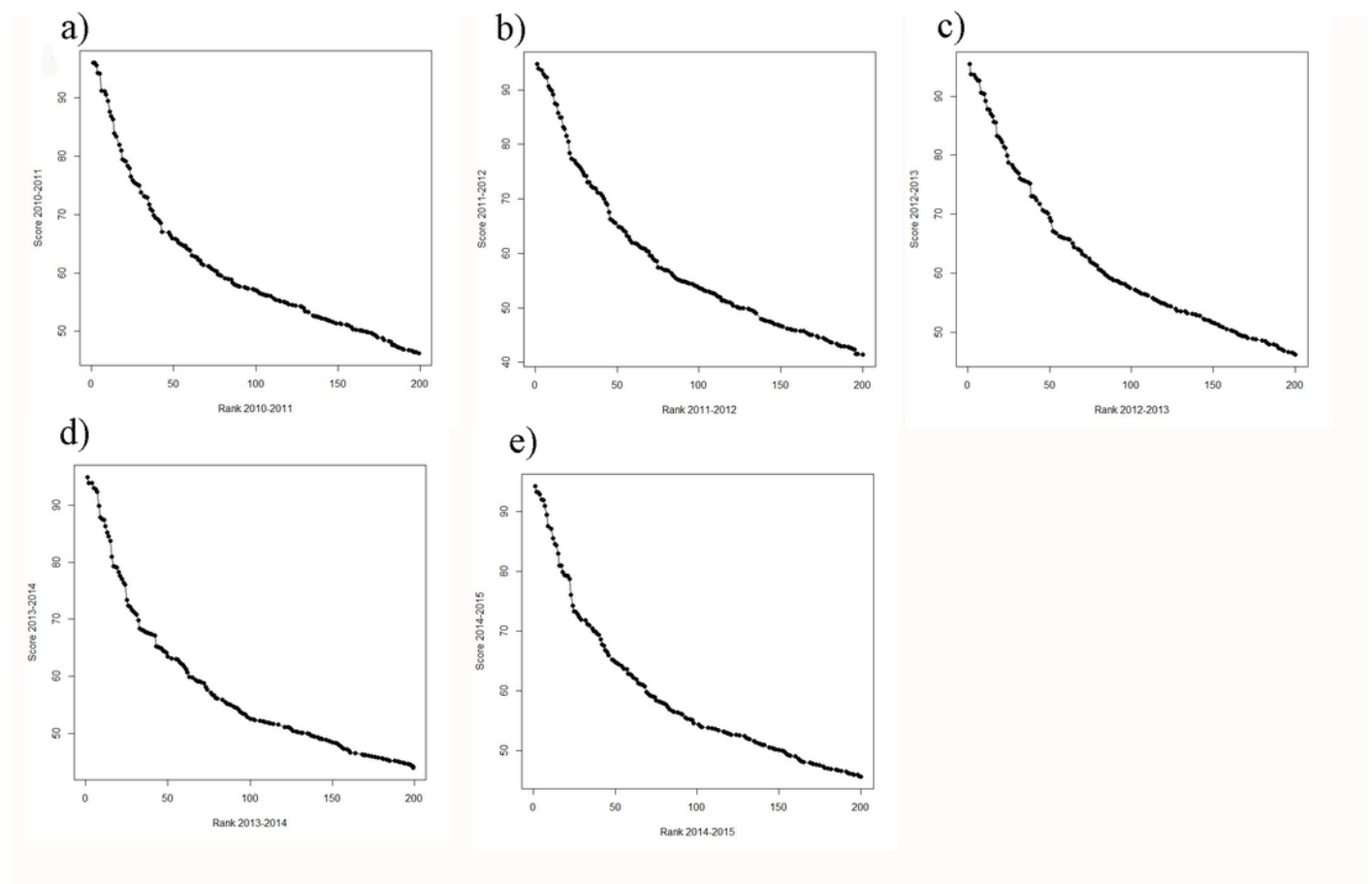



Figure 3

Figure 3) Mean yea-to-year fluctuation (mean of the years 2012, 2013 and 2014) in the THE (%), according to ranking groups. E. g. the percentage of Universities within a ranking group of 50 universities, that changed position beyond the ranking group (moved up, or moved down, respectively dropped out of the ranking).

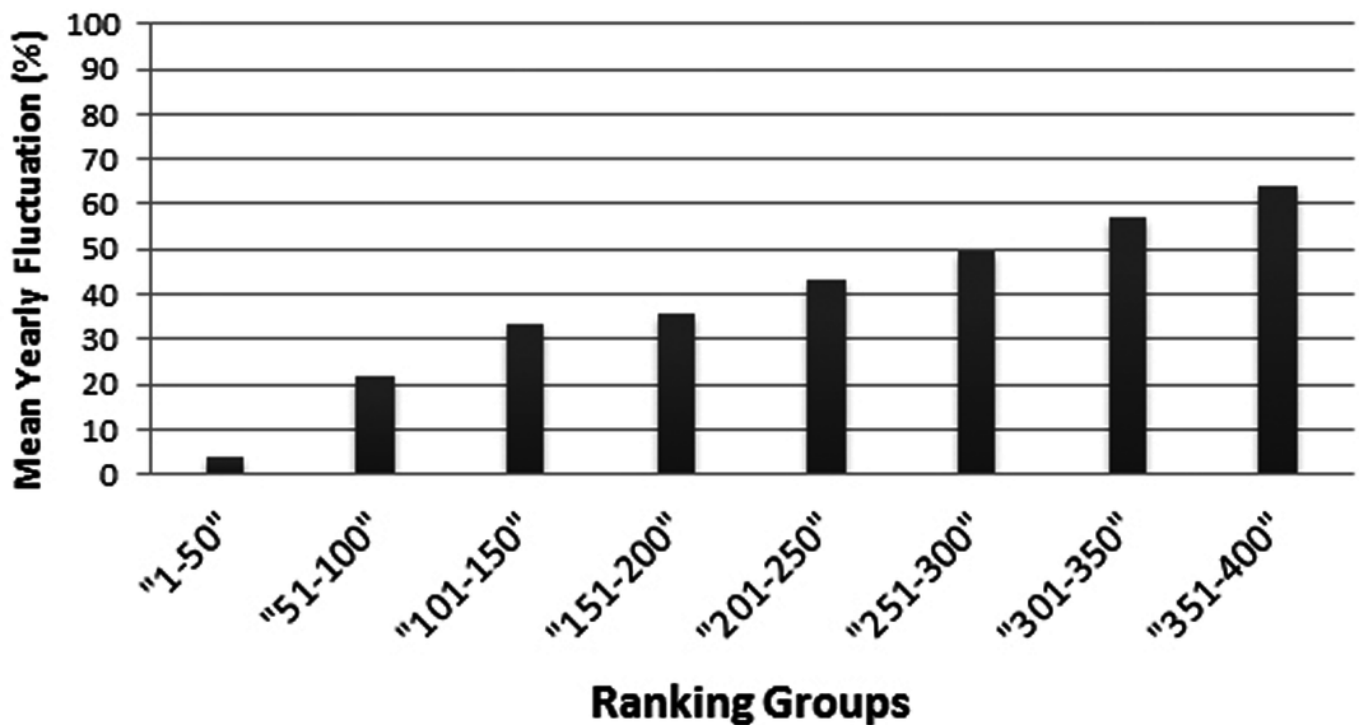



Figure 4

Figure 4 a-d). ARWU: Scores of the year t regressing on the score of the year t-1 from the ranking 2010 on. Figure (4a-4d) Ranks of the year t regressing on the ranks of the year t-1 from the ranking 2010 on (4e-4h). Linear regression line indicates perfect association, e.g. no changes in ranks and scores between two consecutive rankings.

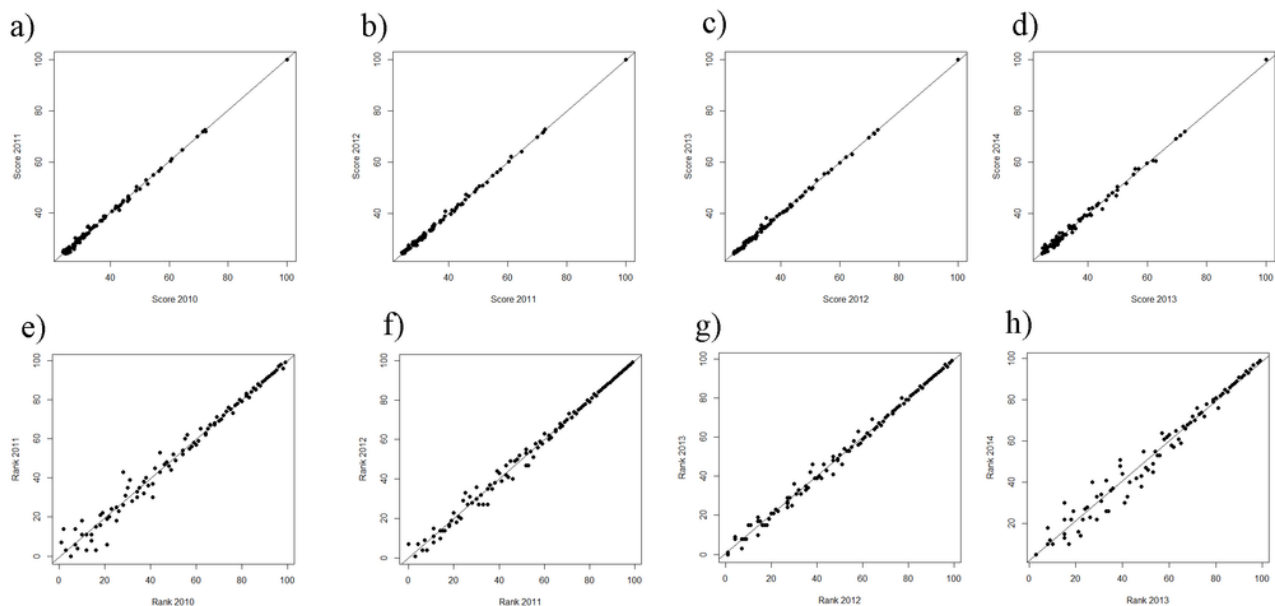



Figure 5

Figure 5 a-e). Scores vs. ranks for the ARWU from 2010 to 2014. a) score 2010 plotted against rank 2010; b) score 2011 plotted against rank 2011 c) score 2012 plotted against rank 2012; d) score 2013 plotted against rank 2013 and e) score 2014 plotted against rank 2014.

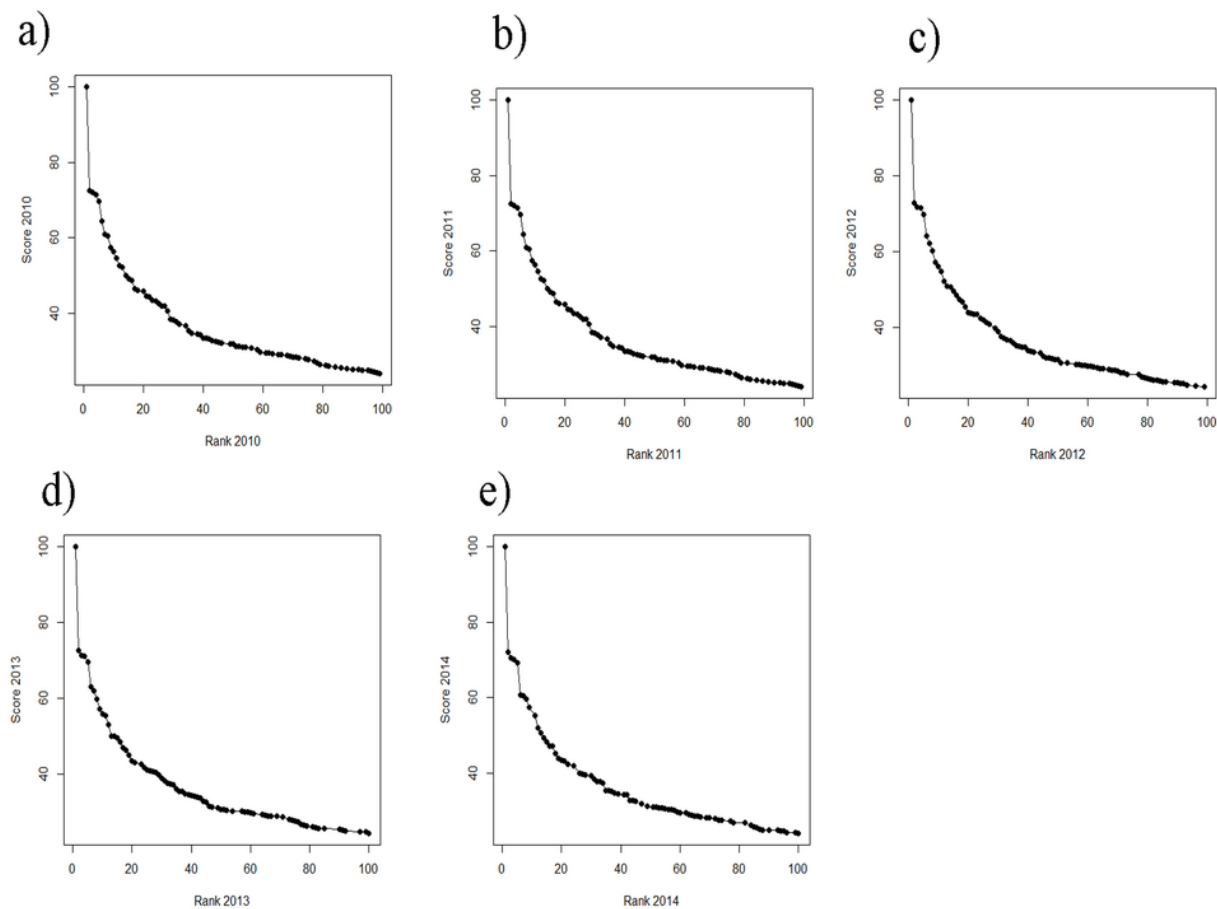



Figure 6

Figure 6) Mean year-to-year fluctuation (mean of the years 2012, 2013 and 2014) in the ARWU (%), according to ranking groups. E. g. the percentage of Universities within a ranking group of 50 universities, that changed position beyond the ranking group (moved up, or moved down, respectively dropped out of the ranking).

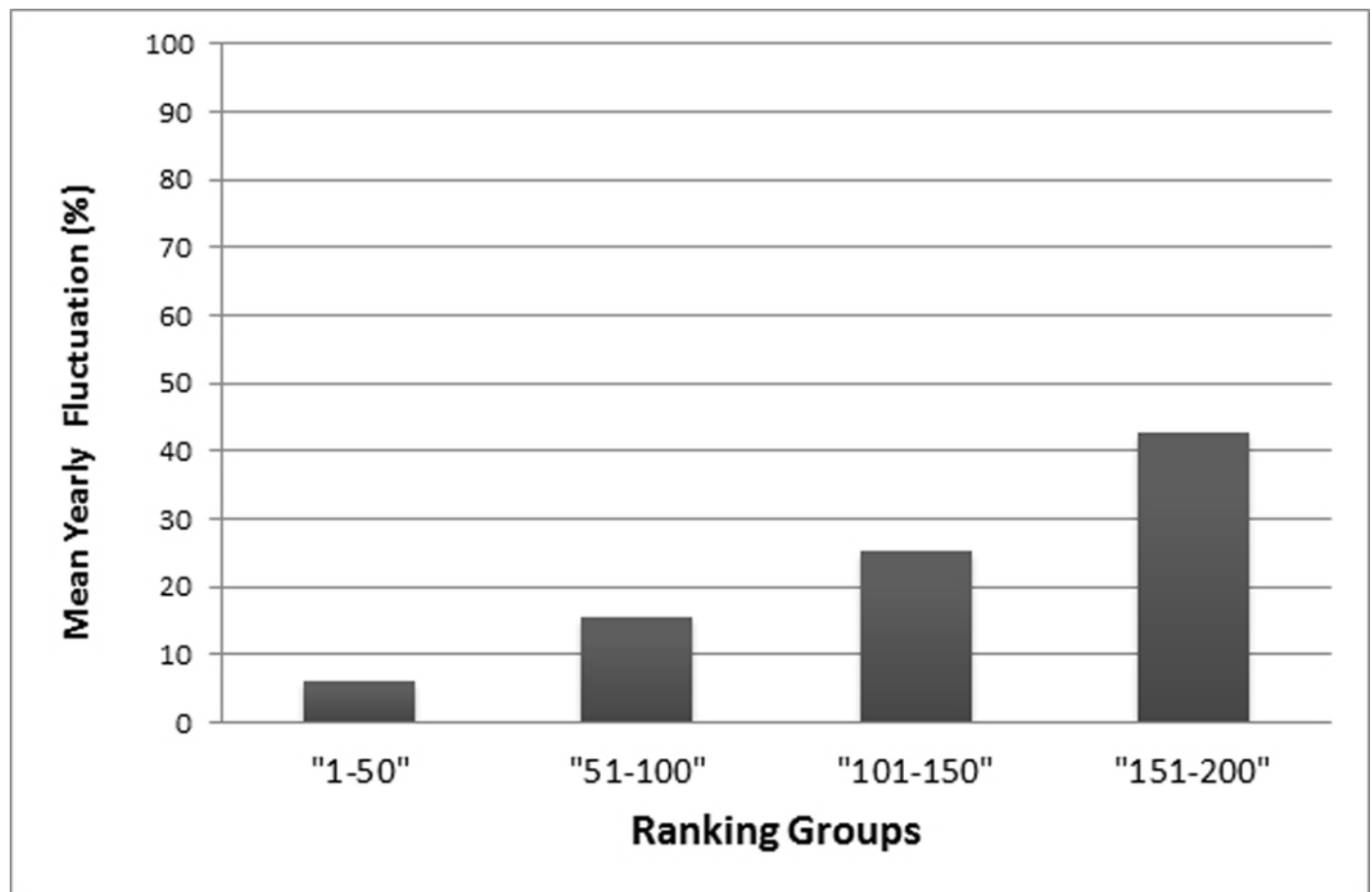



Figure 7

Figure 7) Regression of the ranking positions of the first 100 universities in the THE 2014 on the first 100 universities in the ARWU 2014 (n = 71 universities).