MEASURING THE IMPACT OF STUDENT DIVERSITY ON PERFORMANCE WITH CLASSROOM RESPONSE SYSTEMS

Dr. Stephan SCHMUCKER
Faculty of Economics, Business and Social Sciences
University of Hamburg
Germany
Stephan.Schmucker@wiso.uni-hamburg.de

Dr. Sönke HÄSELER
Researcher
Germany
kontakt@soenkehaeseler.de

Abstract: The diversity of undergraduate students within a given lecture is on the increase – both in terms of their personal traits and their performance. The latter development presents a challenge to lecturers, who may have difficulty adapting their teaching methodology because the students' actual performance often materialises only at the end of the term.

Based on practical experience, this paper shows how classroom response systems, using so-called 'clickers', can be employed to meet this challenge in several ways. A brief test of the material already taught in the first few lectures can provide the lecturer with a rough impression of the performance level in the class right at the start of the term. Yet, more interesting information is to be obtained from a simultaneous survey of potentially performance-related characteristics of the students. Simple statistical analysis will then reveal whether and, if so, which characteristics actually drive student performance. In the best case, the insights thus gained can be used to adapt teaching styles. The paper further argues that short response periods should suffice to elicit meaningful information from the students, meaning that the procedures discussed here do not consume too much lecture time.

INTRODUCTION

Classroom response devices, so-called clickers (see, e.g., Kundisch et al. 2013), are experiencing ever wider use and have been associated with a number of benefits in teaching (Kay/LeSage 2009, Caldwell 2007, Simpson/Oliver 2007, Schmucker 2015). For example, clickers enable a lecturer to conduct single or multiple choice tests, whose questions and corresponding answer choices will be made visible for all students. The students use the clickers to select and to transmit their choices. All responses are recorded and – later on – analysed. Based on the example of a practical application, this paper will show how such analysis can shed light on student diversity, performance, and the relationship between these two phenomena. In doing so, the paper will thus demonstrate a further benefit of the use of clickers, which has so far received little attention.

This agenda is to be seen against the background that the diversity of student bodies is increasing – certainly in Germany (Willich et al. 2011, DSW 2014), but presumably also in many other countries. From this development springs the concern that in large classes it may become increasingly difficult to pursue a teaching style that does justice to most if not all students (Krüger-Basener et al. 2013, Wielepp 2013). This is because we may assume that certain characteristics of the students correlate with their performance and thus, increasing diversity in terms of those characteristics may be expected to entail an increasing divergence of performance.

Lecturers who encounter a new class would well like to have some reliable data on the composition of the student body so as to be able to adapt their teaching styles. Relevant information might for example include the level of performance, prior education (e.g. A-levels or equivalent), or language proficiency. Such information will typically not be available, or at best in unsatisfactory quality, such as might be gained from simply eyeballing the group of students. Information on student performance is altogether lacking, and any attempt to predict performance based on visible student characteristics is prone to error and prejudice. The desired information will usually only materialise during the term (from oral participation) or even at the end of it (from exams) – by which time it is too late to make any use of it.

For these reasons, the authors wish to propose the following procedure: Right at the start of a term in which the lecturer takes on a new class and as soon as sufficient material has been taught for a short test, such a test is conducted using clickers. The test of the actual teaching contents is preceded by a set of questions designed to retrieve information on student characteristics which the lecturer suspects may correlate with performance. Within hours of the test, the lecturer will thus be able to generate the following insights about the class:

1) **DESCRIPTION OF THE STUDENT BODY.** How diverse is the class with respect to the chosen sociodemographic traits? Simple descriptive statistics can already be quite informative, e.g. if the lecturer learns that only 10% of students in a quantitative methods course specialised in mathematics in their prior education.
2) **APPAISING THE LEVEL OF PERFORMANCE.** Primary interest is on the number of lecture-related questions correctly answered – both in terms of the average and the level of dispersion. A high average and low dispersion (the best case) will suggest a different teaching style than a low average and high dispersion (worst case). A certain level of teaching experience in the lecture course in question is required, though, to correctly assess the results.
3) **RELATIONSHIP BETWEEN SOCIODEMOGRAPHIC FACTORS AND PERFORMANCE.** It might be of interest, for example, to learn (from '1' above) that for two-thirds of the class, the language of instruction is not their native language. This could be interpreted as a problem. Statistical analysis may show, however, that performance is actually independent of language skills.

A lecturer who possesses valuable information in these three fields already at the start of the term will be able to adapt her teaching accordingly and, thus, to achieve better learning outcomes. Using the example of trials conducted in the class *Einführung Personal* (Introduction to Human Resource Management) at the University of Hamburg during the winter term 2013/2014, we shall demonstrate how clickers can be employed to collect such information.

Besides providing some exemplary answers in the three areas of interest, this article will argue that the collection of the underlying information is practical and suitable for widespread application in the sense that it requires only a small amount of lecture time: A few minutes of a single lecture suffice to retrieve meaningful data.

DIVERSITY

Four sociodemographic characteristics were selected for the present study at the University of Hamburg. In practical terms, prior to answering the set of questions that relate to the lecture content, the students were presented with four questions concerning their personal traits, and they were asked to transmit the answers (in terms of categories, e.g. age bracket) via the clickers. In other contexts, depending on the lecturer's aims, other characteristics than the ones described here – and different numbers of them – will be appropriate. In the following, we will elaborate on the backgrounds of these characteristics and present some simple descriptive statistics of their manifestation within the group of students examined.

The first characteristic concerns the students' gender, which, however, is only of subordinate interest to the present investigation as in this specific context there is in fact little reason to suspect a correlation of gender with the students' performance. The case may, by contrast, be quite different in other contexts. Consider, for example, a study course that is strongly dominated by students of one sex (e.g. naval engineering versus midwifery). One might expect that the members of the minority have a particular motivation to pursue the course and that their performance therefore exceeds that of the majority. It might be worthwhile for the lecturer to anticipate such a potential effect.

While the students' sex is probably the characteristic that is most easily determined visually, its later association with performance is only possible if the data is recorded electronically, i.e. via clickers. In the class examined, 102 out of 121 students responded to all four sociodemographic questions. The group comprised 46 females and 56 males.

Next, the student's age was recorded, resulting in the frequency distribution displayed in Figure 1. Information on this characteristic, too, would be obtainable from a mere visual inspection of the classroom. Analogously, the added value of data collection via clickers also lies in the subsequent statistical association between age and performance. Once again, the course examined provides no reason to suspect any age effect. Yet imagine, for example, teaching the English language at an Eastern European university, where the language was very rarely taught in high schools before the early 1990s. In such a situation, mature students may have significantly lower prior language skills and may thus be disadvantaged in their studies.

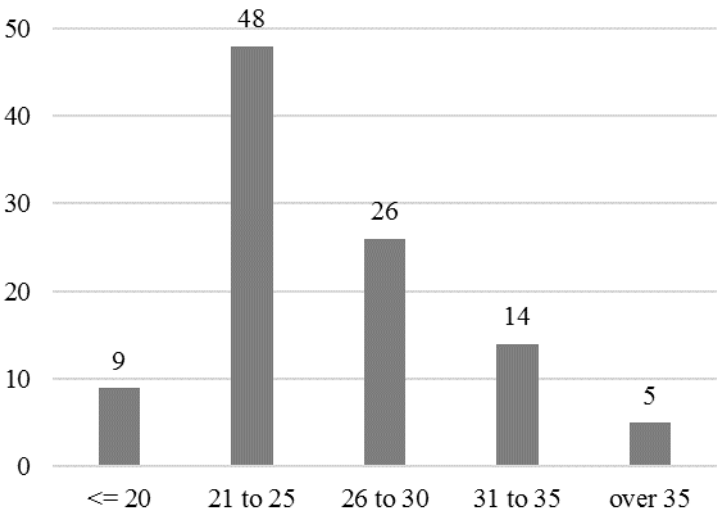

Figure 1. Age

The third characteristic concerns the manner in which the students qualified for entrance to the university. At least in Germany, universities and courses of studies are increasingly opening up to applicants who have not passed the *Abitur* (high school leaving exam equivalent to A-levels) (KMK 2014). It is not far-fetched to suspect an association between the students' prior education and their performance in class (Erdel 2010, Jirjahn 2007). A lecturer may wish to know whether the class she faces for the first time comprises 90% or only 50% students with A-levels. The descriptive statistics for the class examined are shown in Figure 2. In this case, three entrance options besides *Abitur* were distinguished: "*Fachabitur"* (a specialised form of *Abitur* that can be obtained with one year less of education) in conjunction with an oral entrance exam; written and oral "entrance exam"; and "other" (e.g. master craftsmen).

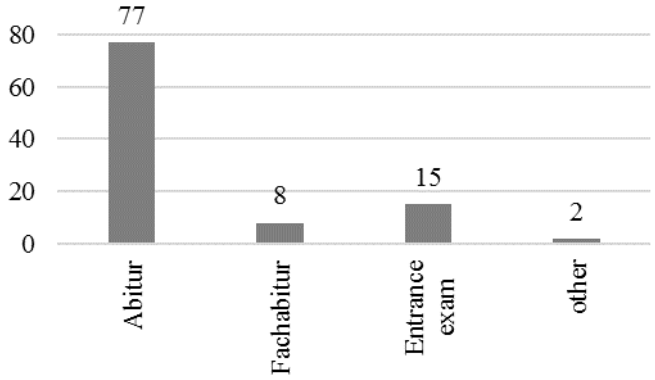

Figure 2. University entrance qualification

Finally, the students were asked about their migratory background. Plausibly, students who were not, or whose parents were not or only partially socialised in the country where the instruction takes place may have greater difficulty getting their bearings in the specific organisational, social and educational environment they find themselves in. A lecturer who ascertains a relatively large share of non-native students in the class and who, in the course of the analysis described in the next section, also learns that such a background can impede the students' academic success, might pay special attention to such students' needs, for example by providing them with additional information to help them navigate their studies. Figure 3 shows the frequency distribution as pertaining to the student body examined, distinguishing between students without any migratory background ("no MB"), students with German citizenship and "1st / 2nd generation migratory background", respectively, and students with a "foreign citizenship".

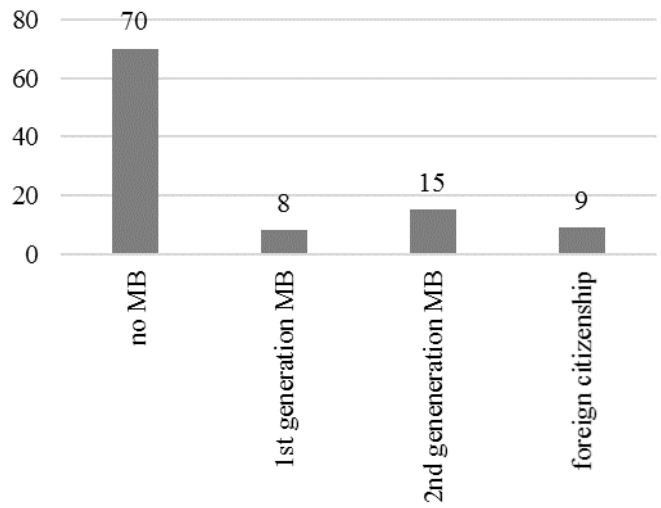

Figure 3. Migratory background (MB)

In other contexts, the lecturer's information needs will suggest the retrieval of different characteristics, including for example:

- prior education, e.g. certain areas of specialisation in high school
- prior professional experience, e.g. having worked as a nurse before studying medicine
- A query of the students' language skills (be it by way of self-assessment of through a short language test using clickers) can help the lecturer decide whether to incorporate foreign-language literature in the course.

PERFORMANCE

In our application at the University of Hamburg, the four sociodemographic questions were followed by a set of eight questions on contents already taught in the introductory HRM course. Each question was accompanied by four to five answer choices, of which the students were to select one using their clickers. The questions and answers are available on request. The ensuing frequency distribution of the number of correct responses is depicted in Figure 4. We do not differentiate between wrong answers and cases in which the students failed to provide any response. The average value is 4.14; the curve is reminiscent of a normal distribution. Measures of dispersion are not informative in this case since we lack the opportunity for comparison – across time or across multiple groups of students.

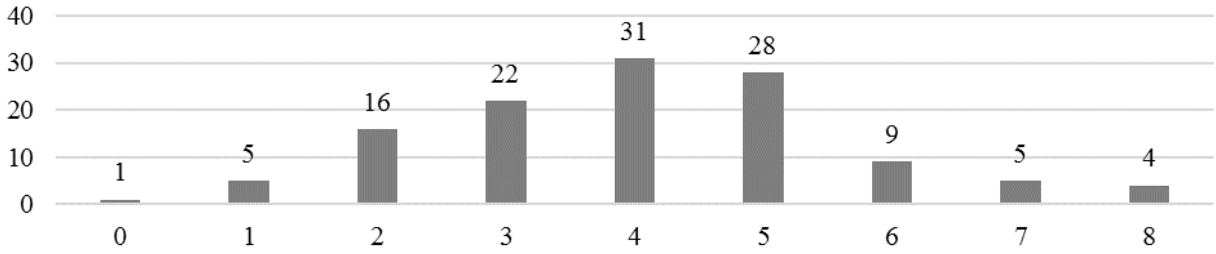

Figure 4. Number of correct responses

In a next step, we relate the students' sociodemographic characteristics to their performance. For this purpose, the number of correct responses becomes the dependent variable in a regression of performance on the four sociodemographic factors. The regression covers the 102 students who responded to all four sociodemographic questions. The answers to each of these four questions forms a categorical or nonmetric variable (age classes, male/female, etc.), which cannot directly be included in a regression. Instead – with the exception of a base (or omitted) category for each variable -, each category is assigned a newly-created dummy variable which can only assume the values of 1 (category applies) and 0 (does not apply). Thus, splitting the four categorical variables into the necessary number of dummy variables, we obtain a total of 11. Table 1 shows the assignment of the

different categories of each of the four characteristics to new dummy variables, where "---" denotes the omitted categories.

Personal traits	**Categories**	**Dummies Model I**	**Dummies Model II**
Gender	male	---	---
	female	female	female
Age	up to 20 years	---	young
	21 to 25 years	25	
	26 to 30 years	30	---
	31 to 35 years	35	
	over 35 years	35+	
University entrance qualification	Abitur	---	Abitur
	Fachabitur	FA	---
	Entrance exam	EE	
	other	other	
Migratory background (MB)	no MB	---	no MB
	MB 1st generation	MB1	---
	MB 2nd generation	MB2	
	foreign citizenship	other	

Table 1.
Transforming the sociodemographic characteristics into dummy variables

In 'Model I' the categories were transformed to dummies one-for-one. To estimate the model, we purposely relied on the simplest OLS procedure, which can be executed in Microsoft Excel, so as not to create any artificial obstacles to imitation by lecturers with only a limited knowledge of statistics. The results are not reproduced here – for the sole reason that no statistically significant effect on performance was found for any of the characteristics or its categories.

In 'Model II', the dummy variables were assigned in such a way that each characteristic now only consists of two categories, e.g. age up to 25 years ("young" = 1) or above ("young" = 0). This assignment was done in light of the regression coefficients obtained in model 1 in such a fashion that the probability of finding significant results was maximised. The estimation results of model II, too, are quickly summarised: The only statistically significant outcome suggest that students without any migratory background marginally outperform those with foreign roots ($z = 2.259$).

At this point, readers may raise the objection that our failure to find performance effects (with one exception) is not due to the actual lack of such effects but rather to the unsophisticated methodology employed, be it with respect to the quality of the sample, the measurement of performance or the estimation strategy. While such an objection could not be altogether rejected, it would, however, miss the point of the present exercise. The point is to provide lecturers with a means of quickly and easily discovering any potential strong relationships between the students' characteristics and their performance so that, in the best case, lecturers may be able to respond to such relationships by adopting their teaching styles. Performance effects that are so weak that they can only be detected with sophisticated methods – as may be the case in the setting described here – are therefore hardly of interest.

Furthermore, it must be noted that this (in a statistical sense) negative result of insignificant performance effects is indeed rather good news for the lecturer, as well as for the university: The observable, pronounced (and arguably increasing) diversity of students does not appear to systematically entail a divergence of performance levels. According to our data at least, the concern that increasing diversity may pose a challenge to teaching is therefore unwarranted.

RESPONSE TIMES

The use of clickers has repeatedly been criticised for its consumption of lecture time (Kay/LeSage 2009, Freeman et al. 2007, Caldwell 2007). This investigation suggests, however, that the provision of relatively brief response times is quite sufficient to achieve meaningful results.

Figure 5 shows the distribution of a total of 856 responses to the eight questions over intervals of ten seconds. Most responses were transmitted after 10 to 20 seconds. The increased frequency in the final interval may be explained as follows: Those students who do not know the correct response but who also hesitate to simply guess will tend to use up all the available time (the students were shown the countdown of time) to search their memories for any clues that might permit an informed response after all. The number of responses that were given in the first ten seconds is quite remarkable. Indeed, 76 responses were even transmitted within the first five seconds. Such extremely short response times are attributable to the fact that when introducing some of the questions, the lecturer permitted a few second to elapse between showing the question with its answer choices and starting the countdown.

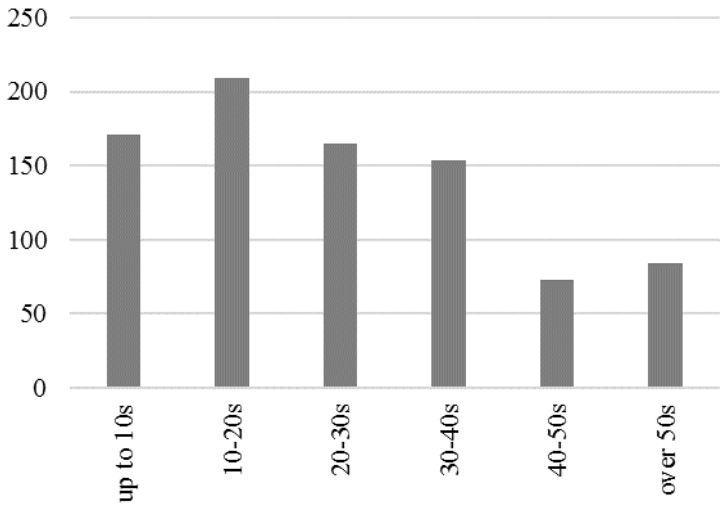

Figure 5. Histogram of response times

Of greater interest though for the purpose of this study is the relationship between the actual response times and the quality of the responses. Comparing the average response time of all correct answers (22.7 seconds) with that of all incorrect answers (28.7 seconds), we find the difference to be highly significant ($t = 5.56$).

This impression is confirmed by Figure 6, which shows the percentage of correct answers in each time interval. The probability of a correct response falls persistently as students take longer to transmit their answers. Once again, the final interval is an exception – and the reason for this could be the same as above: Towards the end of the available time, the students who respond are primarily those who do not know the correct answer but who have at least used the 60 seconds to exclude some choices with the help of what little knowledge they have.

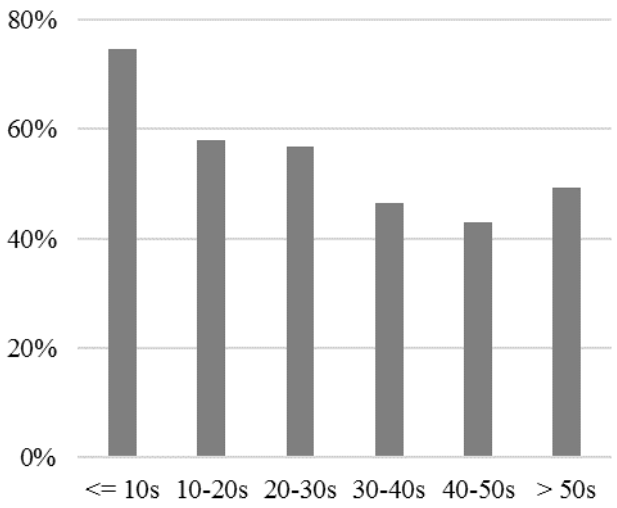

Figure 6. Percentages of correct answers

To support the visual impression of Figure 6 with a statistical test, we ran a probit regression in which a transformation of the (binary) quality of all submitted responses forms the dependent variable. The independent variables consist of the actual response times, their squared values, and seven dummy variables for the lecture-related questions. The detailed results are available upon request and are thus not reproduced here, yet the upshot is quickly summarised: Obtaining highly significant regression coefficients, we find that the likelihood of a correct response falls as response time increases. However, this effect weakens over time and is in fact even reversed for response times in excess of 40 seconds, at which point the probability of a correct response begins to increase again with every additional second that the students take to respond. The regression results thus almost exactly mirror our visual impression.

Yet what does this mean for lecturers who wish to quickly and easily appraise their class? The results suggest that even fairly short periods of available response time suffice for an accurate evaluation of the students' performance. Depending on the length and difficulty of the questions and their associated response options, those who know the correct answer will take no longer than 10 to 20 seconds to read and solve the task. Beyond this time horizon, we likely see an increasing amount of guesswork, which however carries little information value for the lecturer. It ought to be quite possible to conduct a test comprising eight challenging single-choice questions within no more than 15 minutes – including, if desired, a preceding set of sociodemographic questions and the subsequent feedback of the correct responses to the students. Preparing such a test should consume no more than half an hour. With a minimum level of experience, the statistical analysis should take less than two hours. Note that such analysis is in place only once, at the beginning of each term.

CONCLUSION

This article has aimed to provide lecturers with a simple tool to help them appraise a new group of students already at the start of the term and, ideally, to adapt their teaching accordingly. We have shown how clickers, in conjunction with a set of lecture-related questions and a survey of the students' potentially performance-related characteristics, can serve to generate a wealth of valuable information, whose many possible modes of analysis we have only touched upon.

In terms of areas for future research, the influence of response time on performance appears to be the most pressing issue. The present study has merely investigated the association between the quality of the answers and the response time actually needed – while the *available* response time remained constant. It would be interesting to see whether – as we have only been able to presume so far – meaningful results could still be obtained if the available time were reduced to, say, 30 seconds.

ACKNOWLEDGEMENTS

The authors wish to thank the faculty's eLearning Office, represented by Dr. Heiko Witt, for technical assistance with conducting the clicker tests, as well as Jerg Gutmann for valuable advice regarding the statistical analysis.

REFERENCES

Caldwell, J.E. (2007). Clickers in the large classroom: Current research and best-practice tips. Life Sciences Education, 6(1), 9-20.

DSW (2014). Studying today in Germany: 20th social survey of Deutsches Studentenwerk on the social dimension of higher education. Berlin: Heinrich Druck + Medien.

Erdel, B. (2010). Welche Determinanten beeinflussen den Studienerfolg? Eine empirische Analyse zum Studienerfolg der ersten Kohorte der Bachelorstudenten in der Assessmentphase am FB Wirtschaftswissenschaften der Friedrich-Alexander-Universität Erlangen-Nürnberg. Berichte des Lehrstuhls für Soziologie und Empirische Sozialforschung. Bericht 2010-2.

Freeman, M. et al. (2007). Factors affecting educational innovation with in class electronic response systems. Australasian Journal of Educational Technology, 23(2), 149-170.

Jirjahn, U. (2007). Welche Faktoren beeinflussen den Erfolg im wirtschaftswissenschaftlichen Studium? Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, 59(3), 286-313.

Kay, R.H., LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature, Computers & Education, 53, 819-827.

KMK - Kultusministerkonferenz (2014). Synoptische Darstellung der in den Ländern bestehenden Möglichkeiten des Hochschulzugangs für beruflich qualifizierte Bewerber ohne schulische Hochschulzugangsberechtigung auf der Grundlage hochschulrechtlicher Regelungen [On-line]. Available: www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2014/2014_08_00-Synopse-Hochschulzugang-berufl_Qualifizierter.pdf. (02/23/2015)

Krüger-Basener, M. et al. (2013). Heterogenität als Herausforderung für Lehrende der angewandten Technikwissenschaft im Teilprojekt Nord. In: M. Bülow-Schramm (Ed.): Erfolgreich studieren unter Bologna-Bedingungen? Ein empirisches Interventionsprojekt zu hochschuldidaktischer Gestaltung. Bielefeld: Bertelsmann, 162-190.

Kundisch, D. et al. (2013). Classroom Response Systems. Informatik Spektrum, 36(2), 389-393.

Schmucker, S. (2015). Cognitive activation in mass lectures through electronic voting systems in the lecture theatre. In: International Conference on Quality in Higher Education - ICQH 2014, Proceedings Book, Sakarya/Turkey, 67-72.

Simpson, V., Oliver, M. (2007). Electronic voting systems for lectures then and now: A comparison of research and practice. Australasian Journal of Educational Technology, 23(2), 187–208.

Willich, J. et al. (2011). Studienanfänger im Wintersemester 2009/10. Wege zum Studium, Studien- und Hochschulwahl, Situation bei Studienbeginn. HIS: Forum Hochschule 6/2011 [On-line]. Available: www.dzhw.eu/pdf/pub_fh/fh-201106.pdf. (02/23/2015).

Wielepp, F. (2013): Heterogenität. Herausforderung der Hochschulbildung im demografischen Wandel. In: P. Pasternack (Ed.), Jenseits der Metropolen. Leipzig: Akademische Verlagsanstalt, 363-387.

