

Audience-response systems for evaluation of pediatric lectures – comparison with a classic end-of-term online-based evaluation

Abstract

Aim: Course evaluations are often conducted and analyzed well after the course has taken place. By using a digital audience response system (ARS), it is possible to collect, view and discuss feedback during or directly following a course or lecture session. This paper analyzes a student evaluation of a lecture course with ARS to determine if significant differences exist between the results of the ARS lecture evaluation and those of the online evaluation at the end of the semester. In terms of the overall evaluation, consideration is given to the level of students' prior knowledge, the presentation of the lecture material by the lecturers and the relevance of the lecture topic for students.

Method: During the 2011-12 winter semester, the lecture on Pediatrics at the Freiburg Center for Pediatric and Adolescent Medicine (*Zentrum für Kinder- und Jugendmedizin (ZKJ) Freiburg*) was evaluated using ARS. Thirty-four lectures were evaluated by an average of 22 (range 8-44) students, who responded to four questions each time an evaluation took place.

Results: On a 6-point Likert scale (1=very good to 6=deficient), the students rated their level of preparedness with a mean of 3.18, the presentation of the lecture with 2.44, and the relevance of the lecture topic with 2.19. The overall evaluation of the lecture course by means of ARS resulted in 2.31. The online evaluation conducted at the end of the semester yielded a score of 2.45. Highly significant correlations were seen between the results of the ARS for the overall evaluation, assessment of prior knowledge, lecture presentation, and the estimated relevance of the lecture topic.

Conclusion: The use of ARS is suitable for immediate evaluation of lectures, in particular regarding timely feedback for the individual lecturer-lecturers. In comparison with an end-of-term evaluation, ARS yielded a better assessment.

Keywords: Audience Response System, ARS, Evaluation, Pediatrics, Lecture

Sebastian Felix
Nepomuk Bode¹
Christine Straub¹
Marianne Giesler²
Silke Biller³
Johannes Forster⁴
Marcus Krüger¹

1 Universitätsklinikum
Freiburg, Zentrum für Kinder-
und Jugendmedizin, Freiburg,
Germany

2 Universität Freiburg,
Kompetenzzentrum
Evaluation Baden-
Württemberg, Freiburg,
Germany

3 Universität Basel,
Medizinische Fakultät,
Studiendekanat, Basel,
Switzerland

4 St. Josefskrankenhaus,
Freiburg, Germany

1. Background

In Germany the medical licensing regulations (§ 2 subs. 9 ÄAppO) stipulate that educational courses must be regularly evaluated. The state law governing higher education in Baden Württemberg (§ 5 subs. 2) provides for student participation. Student evaluation of university courses is well established at medical schools and is considered reliable under certain conditions [1], [2]. Student evaluations are viewed as reliable particularly when they focus on individual courses and the lecturers receive direct feedback, for instance within the context of an lecturer coaching session [3], [4]. Both the delayed evaluation at semester's end and the general evaluation of a lecture course only offer limited assistance for improving teaching, especially if, as is the case at the

Freiburg Center for Pediatric and Adolescent Medicine (ZKJ), different areas of Pediatrics are covered by different lecturers, or the same ones are covered by different lecturers. The question arises to what extent any evaluation results gathered under these conditions are valid, and if they can be used to improve the lecture course.

The goal is to have an evaluation instrument that, with reasonable effort on the part of lecturers and students, can provide an lecturer with reliable and valid information in a timely manner concerning the quality of his or her teaching and indicate any need for improvement. Audience response systems (ARS) can provide timely information on the quality of courses, and targeted use of this still new evaluation instrument can at least partially counteract student weariness regarding evaluations [5]. The use of ARS in lecture sessions leads to higher levels

of student participation and alertness [5], [6], [7], [8]. However, there is little data currently available on this form of course evaluation [7], [9], [10]. Lecturers see the rapid availability of evaluation results as the primary advantage [7], [9]. One effect of ARS, which must be viewed critically, is that the evaluation of lectures and lecturers is more positive when compared to other forms of evaluation [9], [10], [11], [12].

2. Aim of the study

Previously, the lecture in Pediatrics at the ZKJ was evaluated at the end of the semester using an online-based tool. A individual critique of the teaching by different lecturers did not take place. The lecturers criticized this lack of direct feedback on their own lectures in regard to content and teaching.

The aim of this investigation was to compare the results of the end-of-semester evaluation with the overall evaluation conducted with ARS. Using the variables “student preparation”, “presentation of lecture”, and “relevance of the lecture topic to students”, factors having a potential influence on the evaluation results were also to be investigated.

3. Method

3.1 Procedure

During the 2011-12 winter semester all of the lectures on Pediatrics at the ZKJ were evaluated using ARS. All of the attending students were given a device known as a “clicker” at the beginning of the lecture session. No verification was done to ascertain if all students actually did respond. The evaluation took place immediately after the lecture using the PowerVote® ARS voting system (La Générale Multimédia, Clichy, France). Four questions were posed:

1. At the start of this lecture, I was already prepared for the topic;
2. The way in which the lecture is presented (language, media, speed) helps me understand the content;
3. The lecture content is probably relevant for later professional practice (even for non-pediatricians);
4. My overall evaluation of the lecture is....

Two questions addressed the quality of teaching (questions 2 and 4). Two questions were selected to gauge the extent to which the self-assessed relevance of the topic and the self-assessed level of advance preparation correlate with the presentation and overall assessment of the lecture (questions 1 and 3).

The evaluation was done using a 6-point Likert scale (1=very good to 6=deficient). By asking these questions, the lecturers were to be given direct feedback at the end of each lecture session. At the end of the semester all of the lecturers received not only the results for their own

lectures, but also the mean values of all evaluations for comparison.

Since lecturers often attribute – primarily negative – student evaluations of their teaching to factors outside their influence, we looked to see if the overall assessment (question 4) correlated with student preparation (question 1) and the relevance of the lecture (question 3). In addition, the variable of content presentation was included as a control variable. The evaluations of this variable have to be more closely connected to the overall assessment than the evaluations of the other variables [13]. The score for the overall assessment from the ARS was compared with the score from the online, end-of-semester evaluation, which was conducted four months after the end of the first block in Pediatrics. We postulated that, as shown in prior studies [9], [10], [11], [12], a better evaluation result would be seen for the ARS when compared to the end-of-term evaluation.

3.2 Statistical analysis

The ARS evaluation was analyzed with IBM SPSS Version 20 (IBM SPSS Statistics for Windows, version 20.0. Armonk, NY: IBM Corp.). The lecturers were assigned numerical codes to guarantee anonymity during analysis and dissemination of the results. Correlations were calculated based on Pearson, and t-tests were carried out. To determine the effect size of the mean differences, Cohen’s *d* was calculated. Due to unavailable raw data for the end-of-term evaluation, the mean values of the ARS overall assessment and the mean values for the end-of-term evaluation were compared using a one-sample t-test.

4. Results

The lecture topics covered the areas of specialized general pediatrics, pediatric cardiology, pediatric neurology, and pediatric hematology and oncology. During the 2011-12 winter semester, the lectures on Pediatrics at the ZKJ were held in two instructional blocks of 17 lectures each. The titles of the lectures and the material covered were identical for both blocks, as was the four-week span over which the lectures took place. Eighty students in their 8th semester of study participated in each block. Attendance was not compulsory and the number of participants was not counted. Twenty-one lecturers took part in the lectures.

A total of 833 “responses” were submitted by means of ARS. For each question, the mean number of “responses” given per lecture was 22 (SD=7.98; min. 8 - max. 44). The results for the four questions are presented in Table 1. In Block 1, 343 - 370 “responses” were given per question; in Block 2 the number was 405 - 463. Ninety-seven students (61% response rate) evaluated the 2011-12 winter semester pediatric lecture course using the end-of-semester evaluation online. There is a significant difference between the overall evaluation using ARS and

Table 1: Means and standard deviation of the evaluation using ARS and mean of the online, end-of-term evaluation (standard deviation not available).

Question	Mean	Standard deviation
1: Preparation	3.18	± 1.24
2: Lecture presentation	2.44	± 1.16
3: Relevance	2.19	± 1.07
4: Overall evaluation	2.31	± 1.18
End-of-semester evaluation	2.45	--

the overall evaluation of the end-of-semester evaluation ($M_{ARS}=2.31$, $SD=1.178$; $M_{SemEnd}=2.45$; t value -3.557 , df 787 , $p<.0001$).

The results of the ARS evaluations for both lecture blocks are illustrated in Table 2. For all of the questions there are significant mean differences, with the effect sizes (Cohen's d) lying in the lower to middle range. The results for Block 2 concerning lecture presentation are less favorable than those for Block 1. The students who attended the Block 2 lectures reported that they were less prepared in comparison with the students in Block 1. Furthermore, they estimated the relevance of the course topic to be less important than their peers in Block 1. Since it was very striking that in both blocks the range of ratings varied for the individual lecturers regarding lecture presentation (Block 1: $M=1.46$ to $M=3.10$; Block 2: $M=2.00$ to $M=4.10$) and the overall evaluation (Block 1: $M=1.33$ to $M=3.38$, Block 2: $M=1.91$ to $M=4.43$), additional analyses (not presented here) were performed separately for each block using the factor "lecturer". The results of these analyses revealed that the mean evaluations in Block 2 for the two most poorly rated lecturers differed significantly from the mean values for the evaluations of all the other lecturers, while in Block 1 significant evaluation differences were more frequently seen between the lecturers in this same lower range. This means that the evaluations of the lecturers in Block 1 are better on the whole, but more heterogeneous. The results of the intercorrelations between the evaluations gathered through ARS are presented in Table 3. The assessments regarding lecture presentation and the topic's relevance correlate most strongly with the overall evaluation of the course. The lecture presentation and course relevance correlate with each other in a highly significant manner.

5. Discussion

5.1 ARS participation

The use of ARS in the lecture setting is rated positively by both students and teachers [5], [6], [7], [8], [9]. No statement about the acceptance of the system can be made based on this study, since this aspect was not specifically investigated. However, oral feedback from students and lecturers was consistently positive. The ratio of students evaluating via ARS to all students present was not documented. However, all students did receive a clicker and were requested multiple times during the course to make use of it. In addition, the clicker was used as a teaching tool during the lectures.

For courses not requiring attendance, an absence rate of 18.5-70% is described [14], [15], [16]. If the number of votes cast is taken to be the actual number of students present, then the average rate of attendance (with a mean of 22 votes per question) is in the lower range of the percentages reported in the relevant studies. Should this study be repeated, documenting the number of students present will be necessary for better assessment of the results.

5.2 Preparation, Presentation & Relevance

An lecturer's teaching and lecture style can positively or negatively affect the overall evaluation of a course, even if only to a small degree [13], [17], [18]. Our data also show that, above all, the presentation of the material is closely connected to the overall evaluation – even more strongly than the relevance of the topic itself. Although advance preparation on the part of students also correlates significantly with the overall evaluation, this connection is slight in our study. It is assumed that the lectures which students find relevant will be evaluated more positively [19]. Since this perception is not directly connected with the actual course, it is also presumed that a bias

Table 2: Comparison of the evaluation results for Block 1 (343≤N≤370) and Block 2 (405≤N≤463). The same lectures were offered in both blocks.

Question	Block	Mean	Standard deviation	Significance p	T-test T
1: Preparation	1	3.07	+/- 1.23	.000	-3.53
	2	3.38	+/- 1.24		
2 : Lecture presentation	1	2.08	+/- 0.99	.000	-5.4
	2	2.51	+/- 1.24		
3: Relevance	1	2.01	+/- 0.98	.002	-3.18
	2	2.24	+/- 1.14		
4: Overall evaluation	1	2.05	+/- 0.97	.000	-5.54
	2	2.49	+/- 1.28		

**Table 3: Results of the correlation analysis (based on Pearson)
Number of responses per comparison: 541-810.**

Correlation Coefficients			
	2: Lecture presentation	3: Relevance	4: Overall evaluation
1: Preparation	.180**	.147**	.103*
2: Lecture presentation		.631**	.836**
3: Relevance			.612**

** : p = .01 (two-sided); * : p = .05 (two-sided).

variable is involved [20]. In contrast, the presentation of a lecture is primarily dependent on the lecturer [21].

5.3 Observations over time

In the final weeks of 2011, a significantly lower evaluation of the course was seen in Block 2. Two lectures were particularly criticized by the students in this block. It is assumed that the evaluations of these two lectures are primarily responsible for the discrepancy seen between the two blocks. With ARS, poorly rated lectures can be identified during the lecture phase and action can be taken in a very timely manner, for instance in the form of lecturer coaching or revision of the pertinent lecture content. With an end-of-term evaluation, this kind of action is much more limited.

5.4 Overall evaluation, including a comparison with the end-of-semester evaluation

The lecture course on Pediatrics during the 2011-12 winter semester at the University of Freiburg Medical Center (Universitätsklinikum Freiburg) was evaluated on the end-of-semester evaluation with the score of 2.45. The same course received a score of 2.31 on the ARS evaluation. There is a statistically significant difference between these evaluation results. The inherent effect of ARS to lead to a more positive evaluation may be one of the decisive factors determining this outcome [9], [10], [11], [12].

5.5 Feedback for the lecturers

Following the evaluation, the results were presented to the lecturers in an anonymized form. The lecturers could identify the results for their particular lecture based on an assigned number known only to them. The lectures were then subject to an internal review and revised as needed. All of the lecturers were offered training in lecture presentation. A repeat evaluation following these changes has yet to be done; the initial responses of the lecturers and students are positive. Overall, the feedback from lecturers regarding the rapid availability of the evaluation results was very positive. ARS was judged to be a suitable evaluation medium.

6. Conclusion

With little effort, lectures can be evaluated with ARS while they are taking place or immediately afterward. In addition, ARS can be used to make lectures interactive [5]. An important advantage of using ARS to evaluate educational courses is the immediate access to lecturer-specific feedback. This study was able to demonstrate that, by using ARS as an evaluation medium during or after a course, the variables of student preparation, lecture presentation, and relevance of the topic correlate significantly with the overall evaluation of a course.

Individual lectures that are given poor evaluations can lead to a lower evaluation for the course overall. However, this situation is immediately visible in the ARS results, so that timely intervention can take place, for instance in the form of special coaching for the affected lecturer. When relying solely on end-of-term evaluations, this is not possible.

A limitation of this study is that the number of students present in comparison with the number of students actually participating in the ARS evaluation was not determined.

Due to the advantages listed above, evaluations using ARS should be made available to other subjects following further standardization and refinement.

Acknowledgement

We wish to thank all of the lecturers at the *Zentrum für Kinder- und Jugendmedizin Freiburg* and all of the participating students.

Competing interests

The authors declare that they have no competing interests.

References

1. Kuhnigk O, Weidtmann K, Anders S, Huneke B, Santer R, Harendza S. Lectures based on cardinal symptoms in undergraduate medicine - effects of evaluation-based interventions on teaching large groups. *GMS Z Med Ausbild.* 2011;28(1):Doc15. DOI: 10.3205/zma000727
2. Diehl JM. Normierung zweier Fragebögen zur studentischen Beurteilung von Vorlesungen und Seminaren. *Psychol Erz Unterr.* 2003;50:15.
3. Dresele M, Rindermann H, Tinsner K. Consulting of teachers on the basis of course evaluation of students. 1 ed. Lengerich: Pabst Science Publishers; 2007.
4. Rindermann H. Quality of instruction improved by evaluation and consultation of instructors. *Int J Acad Develop.* 2007;12(2):13. DOI: 10.1080/13601440701604849
5. Caldwell JE. Clickers in the large classroom: current research and best-practice tips. *CBE Life Sci Educ.* 2007;6(1):9-20. DOI: 10.1187/cbe.06-12-0205
6. Rahman A, Jacker-Guhr S, Staufenbiel I, Meyer K, Zupanec M, Hahnemann M, Lührs AK, Eberhard J. Anwendung von elaborem Feedback und einem Audience-Response-System in der Zahnmedizinischen Ausbildung. *GMS Z Med Ausbild.* 2013;30(3):Doc35. DOI: 10.3205/zma000878
7. Banks DA. Audience Response Systems in Higher Education: Applications and Cases. Hershey: Idea Group Pub; 2006. DOI: 10.4018/978-1-59140-947-2
8. Kay RH, LeSage A. Examining the benefits and challenges of using audience response systems: A review of the literature. *Comp Educ.* 2009;53(3):9. DOI: 10.1016/j.compedu.2009.05.001
9. Savage J. Using an Audience Response System for Module Evaluation. In: Cardiff University (Hrsg). *Technology Enhanced Education.* Cardiff: Cardiff University; 2013. S.5.
10. Plischko R. Mediale Gestaltung von Vorlesungen in der medizinischen Ausbildung mit Hilfe eines Audience-Response-Systems. Dissertation. München: Ludwig-Maximilians-Universität München; 2006.
11. Miller RG, Ashar BH, Getz KJ. Evaluation of an audience response system for the continuing education of health professionals. *J Contin Educ Health Prof.* 2003;23(2):109-115. DOI: 10.1002/chp.1340230208
12. Hecht S, Adams WH, Cunningham MA, Lane IF, Howell NE. Student performance and course evaluations before and after use of the Classroom Performance System in a third-year veterinary radiology course. *Vet Radiol Ultrasound.* 2012;54(2):114-121. DOI: 10.1111/vru.12001
13. Marburger DR. Absenteeism and undergraduate exam performance. *J Eco Educ.* 2001;32:99-110. DOI: 10.1080/00220480109595176
14. Friedman P, Rodriguez F, McComb J. Why students do and do not attend class. *Coll Teach.* 2001;49:124-133. DOI: 10.1080/87567555.2001.10844593
15. Moore R. Attendance and performance: How important is it for students to attend class? *J Coll Sci Teach.* 2003;32:367-371.
16. Abrami PC. Educational Seduction. *Rev Educ Res.* 1982;52(3):446-464. DOI: 10.3102/00346543052003446
17. Marsh HW, Ware JE. Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. *J Educ Psychol.* 1982;74:17. DOI: 10.1037/0022-0663.74.1.126

18. Centra JA. Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness. San Francisco: Jossey-Bass Pub; 1993.
19. Kek M, Stow S. What makes students happy? Factors influencing student engagement using student evaluation data. Brisbane: FYE Curriculum Design Symposium; 2009. S.6
20. Rindermann H. Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen. Landau: Empirische Pädagogik; 1996.
21. Marsh HW. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. J Educ Psychol. 1984;76(5):54. DOI: 10.1037/0022-0663.76.5.707

Please cite as

Bode SF, Straub C, Giesler M, Biller S, Forster J, Krüger M. Audience-response systems for evaluation of pediatric lectures – comparison with a classic end-of-term online-based evaluation. GMS Z Med Ausbild. 2015;32(2):Doc18. DOI: 10.3205/zma000960, URN: urn:nbn:de:0183-zma0009602

This article is freely available from

<http://www.egms.de/en/journals/zma/2015-32/zma000960.shtml>

Received: 2014-08-04

Revised: 2015-02-03

Accepted: 2015-03-20

Published: 2015-05-13

Copyright

©2015 Bode et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Corresponding author:

Prof. Dr. Marcus Krüger
Universitätsklinikum Freiburg, Zentrum für Kinder- und Jugendmedizin, Mathildenstrasse 1, Freiburg, Germany
Marcus.Krueger@uniklinik-freiburg.de

Audience-Response Systeme zur Evaluation von Pädiatrievorlesungen – Vergleich mit einer klassischen onlinebasierten Semesterende-Evaluation

Zusammenfassung

Zielsetzung: Lehrveranstaltungsevaluationen werden oft zeitverzögert zur Lehrveranstaltung durchgeführt und ausgewertet. Mit dem elektronischen Evaluationsmedium Audience Response System (ARS) kann unmittelbar während oder nach Ende einer Lehrveranstaltung Feedback eingeholt, dargestellt und diskutiert werden. In dieser Arbeit wird eine studentische ARS-Vorlesungsevaluation analysiert. Es wird überprüft, ob signifikante Unterschiede zwischen den Ergebnissen der ARS-Vorlesungsevaluation und der Online-Evaluation am Ende des Semesters bestehen. Hierbei wird der Zusammenhang des studentischen Vorwissens, der Gestaltung der Lehrveranstaltung durch die Lehrenden sowie der Relevanz des Vorlesungsthemas für die Studierenden in Bezug auf die Gesamtbewertung betrachtet.

Methodik: Im Wintersemester 2011/2012 wurde die Pädiatrievorlesung am Zentrum für Kinder- und Jugendmedizin (ZKJ) Freiburg mit Hilfe eines ARS evaluiert. Es wurden 34 Vorlesungen von durchschnittlich 22 (range 8-44) Studierenden mit jeweils vier Fragen evaluiert.

Ergebnisse: Auf einer 6-stufigen Likert-Skala (1=sehr gut bis 6=ungenügend) bewerteten die Studierenden ihr Vorwissen im Mittel mit 3,18, die Vorlesungsgestaltung mit 2,44 und die Relevanz des Vorlesungsthemas mit 2,19. Die Gesamtbewertung der Vorlesung mittels ARS-Evaluation ergab die Note 2,31. In der Online Evaluation am Ende des Semesters wurde die Note 2,45 ermittelt. Es zeigt sich hoch signifikante Zusammenhänge zwischen den durch den Einsatz des ARS erhobenen Ergebnissen der Gesamtbewertung, dem Vorwissen der Studierenden, der Vorlesungsgestaltung sowie der selbsteingeschätzten Relevanz des Vorlesungsthemas.

Schlussfolgerung: Der Einsatz von ARS ist zur unmittelbaren Vorlesungsevaluation, insbesondere im Hinblick auf zeitnahe (lehr-)personenbezogene Rückmeldungen geeignet. Im Vergleich zu einer Semesterende-Evaluation zeigt sich in der ARS-Evaluation eine bessere Bewertung.

Schlüsselwörter: Audience Response System, ARS, Evaluation, Pädiatrie, Vorlesung

1. Hintergrund

In Deutschland gibt die ärztliche Approbationsordnung (ÄAppO: § 2, Absatz 9) vor, dass Lehrveranstaltungen regelmäßig zu evaluieren sind. Das Landeshochschulgesetz von Baden Württemberg (§5, Abs. 2) sieht die Beteiligung der Studierenden vor. Die Evaluation von Lehrveranstaltungen durch Studierende ist an den Medizinischen Fakultäten etabliert. Sie gilt unter bestimmten Voraussetzungen als reliabel [1], [2]. Als zuverlässig gilt die studentische Evaluation insbesondere dann, wenn sie detailliert auf die einzelne Lehrveranstaltung bezogen ist und die Lehrenden ein unmittelbares Feedback erhalten, zum

Beispiel im Rahmen eines kollegialen Coachings [3], [4]. Sowohl die zeitversetzte Evaluation am Semesterende, als auch die Gesamtevaluation eines Vorlesungsblockes, sind in diesem Sinne nur bedingt hilfreich für eine Verbesserung der Lehre, insbesondere wenn – wie auch am Zentrum für Kinder- und Jugendmedizin (ZKJ) Freiburg – im Vorlesungsblock verschiedene Themengebiete von verschiedenen Lehrenden gelesen oder gleiche Themengebiete von unterschiedlichen Lehrenden vorgetragen werden. Somit stellt sich die Frage, inwieweit Evaluationsergebnisse, die unter diesen Bedingungen erhoben wurden, valide sind und ob sie zur Verbesserung der Lehrveranstaltung verwendet werden können. Ziel ist ein Evaluations-Instrument zur Verfügung zu haben, das mit vertretbarem Aufwand für Lehrende und

Sebastian Felix
Nepomuk Bode¹
Christine Straub¹
Marianne Giesler²
Silke Biller³
Johannes Forster⁴
Marcus Krüger¹

1 Universitätsklinikum
Freiburg, Zentrum für Kinder-
und Jugendmedizin, Freiburg,
Deutschland

2 Universität Freiburg,
Kompetenzzentrum
Evaluation Baden-
Württemberg, Freiburg,
Deutschland

3 Universität Basel,
Medizinische Fakultät,
Studiendekanat, Basel,
Schweiz

4 St. Josefskrankenhaus,
Freiburg, Deutschland

Studierende der jeweiligen Lehrperson zeitnah zuverlässige und valide Hinweise zur Qualität der eigenen Lehre geben und möglichen Verbesserungsbedarf aufzeigen kann. Mittels Audience-Response-Systemen (ARS) können zeitnah Hinweise zur Qualität von Lehrveranstaltungen gegeben werden und durch zielgerichteten Einsatz dieses noch neuen Evaluationsinstrumentes der „Evaluationsmüdigkeit“ Studierender zumindest teilweise entgegen gewirkt werden [5]. Der Einsatz von ARS in Vorlesungen führt zu höherer Partizipation und Aufmerksamkeit der Studierenden [5], [6], [7], [8]. Bisher gibt es jedoch zu dieser Form der Evaluation von Lehrveranstaltungen nur wenige Daten [7], [9], [10]. Lehrende sehen vor allem die rasche Verfügbarkeit der Evaluationsdaten als Vorteil an [7], [9]. Ein kritischer zu sehender Effekt von ARS ist die positivere Bewertung von Vorlesungen und Lehrenden im Vergleich zu anderen Formen der Evaluation [9], [10], [11], [12].

2. Ziel der Untersuchung

Bislang wurde die Pädiatrievorlesung am ZKJ Freiburg am Ende des Semesters mittels eines onlinebasierten Tools evaluiert. Eine Beurteilung der Lehrqualität einzelner Lehrender fand nicht statt. Die Lehrenden bemängelten dieses Fehlen direkter Rückmeldungen zur eigenen Vorlesung in Bezug auf inhaltliche und didaktische Aspekte.

Ziel der vorliegenden Untersuchung war es Ergebnisse der Semesterende-Evaluation mit der Gesamtbewertung einer ARS-Evaluation zu vergleichen. Anhand der Variablen „studentisches Vorwissen“, „Vorlesungsgestaltung“ und „Relevanz des Vorlesungsthemas“ für die Studierenden sollten zusätzlich mögliche Einflussfaktoren auf die Evaluationsergebnisse untersucht werden.

3. Methode

3.1 Ablauf der Untersuchung

Im Wintersemester 2011/2012 wurden alle Pädiatrievorlesungen am Zentrum für Kinder- und Jugendmedizin Freiburg mittels ARS evaluiert. Alle anwesenden Studierenden erhielten zu Vorlesungsbeginn ein Abstimmungsgerät. Eine Überprüfung, ob alle anwesenden Studierenden tatsächlich abstimmten, fand nicht statt. Die Evaluation erfolgte direkt am Ende der Vorlesungen mittels PowerVote® ARS-Abstimmungssystem (La Générale Multimédia, Clichy, Frankreich). Es wurden vier Fragen gestellt:

1. „Zu Beginn dieser Vorlesung war ich bereits auf das Thema vorbereitet.“
2. „Die Art, wie die Vorlesung gestaltet ist (Sprache, Medien, Geschwindigkeit), trägt zum Verständnis des Stoffes bei.“
3. „Die Vorlesungsinhalte sind vermutlich für die spätere Berufspraxis relevant (auch für Nicht-Pädiater).“
4. „Meine Gesamtbewertung der Vorlesung ist: ...“.

Zwei Fragen bezogen sich auf die Bewertung der Lehrqualität (Frage 2 und Frage 4). Zwei Fragen wurden ausgewählt, um abzuschätzen, inwieweit die selbsteingeschätzte Relevanz des Themas und das selbst eingeschätzte Vorwissen im Zusammenhang zur Gestaltung und Gesamtbewertung der Vorlesung stehen (Frage 1 und Frage 3): Die Bewertung erfolgte auf einer 6-stufigen Likert-Skala (1=sehr gut bis 6=ungenügend). Anhand dieser Fragen sollten die Lehrenden am Ende jeder Vorlesung eine unmittelbare Rückmeldung erhalten. Am Ende des Semesters erhielten alle Lehrenden sowohl die Ergebnisse ihrer Lehrveranstaltung, als auch die Mittelwerte aller Evaluationen zum Vergleich.

Da die Lehrenden oft die – vor allem negative – studentische Bewertung ihrer Lehre anderen, außerhalb ihres Einflusses stehenden, Variablen zuschreiben, wurde überprüft, ob die Gesamt-Bewertung (Frage 4) mit dem Vorwissen der Studierenden (Frage 1) und der Relevanz der Vorlesung (Frage 3) im Zusammenhang steht. Darüber hinaus wurde die Variable Vorlesungsgestaltung als sogenannte „Kontrollvariable“ mitaufgenommen. Die Bewertungen dieser Variable müsste mit der Gesamtbewertung in einem engeren Zusammenhang stehen als die Bewertungen der anderen Variablen [13]. Die Note der im Rahmen der ARS-Evaluation erhobenen Gesamtbewertung wurde mit der Note der onlinebasierten Semesterende-Evaluation der Vorlesung verglichen, welche vier Monate nach Ende des ersten Pädiatrie-Blocks durchgeführt wurde. Wir postulierten, wie in Voruntersuchungen gezeigt [9], [10], [11], [12], ein besseres Evaluationsergebnis mit der ARS-Evaluation im Vergleich zur Semesterende-Evaluation.

3.2 Statistische Auswertung

Die Auswertung der ARS-Evaluation erfolgte mit IBM SPSS Version 20 (IBM SPSS Statistics für Windows, Version 20.0. Armonk, NY: IBM Corp.). Den Lehrenden wurden Zahlencodes zugeordnet, um deren Anonymität bei der Auswertung sowie für die Ergebnisdarstellung zu wahren. Es wurden Korrelationen nach Pearson berechnet und T-Tests durchgeführt. Zur Bestimmung der Effektstärke der Mittelwertunterschiede wurde Cohens *d* berechnet. Aufgrund nicht zur Verfügung stehender Rohdaten bei der Semesterende-Evaluation wurden die Mittelwerte der ARS-Gesamtbewertung und die Mittelwerte der Semesterende-Evaluation-Bewertung mit Hilfe des „T-Tests bei einer Stichprobe“ miteinander verglichen.

4. Ergebnisse

Die Vorlesungsthemen umfassten die Themengebiete der speziellen allgemeinen Pädiatrie, der pädiatrischen Kardiologie, der Neuropädiatrie und der pädiatrischen Hämatologie und Onkologie. Im WS 2011/12 wurden die Pädiatrie-Vorlesungen am ZKJ Freiburg titel- und inhalts-gleich in zwei Unterrichtsblöcken à 17 Vorlesungen in je 4 Wochen durchgeführt. Je Block nahmen 80 Studierende

im 8. Fachsemester teil. Eine Anwesenheitspflicht für die Vorlesung bestand nicht, eine Überprüfung der Anwesenheit fand nicht statt. 21 Lehrende beteiligten sich an den Vorlesungen.

Mittels ARS-Evaluation wurden insgesamt 833 „Antworten“ abgegeben. Pro Frage wurden im Mittel pro Vorlesung 22 „Antworten“ abgegeben (SD=7,98; Min. 8 - Max. 44). Die Ergebnisse zu den vier gestellten Fragen finden sich in Tabelle 1. In Block 1 wurden pro Frage 343 - 370 "Antworten" abgegeben, in Block 2 405 - 463. Im Rahmen der onlinebasierten Endsemesterevaluation wurde die Pädiatrievorlesung im WS 2011/12 von 97 Studierenden evaluiert (61% Rücklauf). Zwischen der Gesamtbewertung mittels ARS und der Gesamtbewertung der onlinebasierten Semesterende-Evaluation zeigte sich ein signifikanter Unterschied ($M_{ARS}=2,31$, $SD=1,178$; $M_{SemEnd}=2,45$; t-Wert $-3,557$, $df 787$, $p<.0001$).

In Tabelle 2 sind die Ergebnisse der ARS-Evaluationen für beide Vorlesungsblöcke dargestellt. Bei allen Fragen ergeben sich signifikante Mittelwertunterschiede, wobei die Effektstärken (Cohens d) im niedrigen bis mittleren Bereich liegen. Für Block 2 fallen die Ergebnisse für die Vorlesungsgestaltung und Gesamtbewertung weniger günstig aus als für Block 1. Die Studierenden, welche die Vorlesungen in Block 2 besuchten, gaben im Vergleich zu den Studierenden des Blocks 1 an, weniger gut vorbereitet gewesen zu sein. Zudem schätzten sie die Relevanz des Themas der Veranstaltung als weniger wichtig ein als ihre KommilitonInnen in Block 1. Da auffällig war, dass in beiden Blöcken der range der Beurteilungen für die einzelnen Lehrenden im Hinblick auf die Vorlesungsgestaltung (Block 1: $M=1,46$ bis $M=3,10$; Block 2: $M=2,00$ bis $M=4,10$) und der Gesamtbewertung (Block 1: $M=1,33$ bis $M=3,38$, Block 2: $M=1,91$ bis $M=4,43$) variierte, wurden weitere Analysen (hier nicht dargestellt) mit dem Faktor „Lehrende“ getrennt für beide Blöcke durchgeführt. Die Ergebnisse dieser Analysen zeigen, dass sich in Block 2 die mittleren Beurteilungen von zwei Lehrenden, die am schlechtesten bewertet wurden, signifikant von den mittleren Beurteilungen aller anderen Lehrenden unterschieden, während in Block 1 häufiger signifikante Beurteilungsunterschiede zwischen den Lehrenden im oben beschriebenen range festzustellen waren. Dies bedeutet, dass die Beurteilungen der Lehrenden in Block 1 heterogener, jedoch insgesamt besser, waren. Die Ergebnisse der Interkorrelationen zwischen den mittels ARS erhaltenen Bewertungen sind in Tabelle 3 dargestellt. Die Einschätzungen der Vorlesungsgestaltung und die der Relevanz des Themas korrelieren am stärksten mit der Gesamtbewertung der Veranstaltung. Die Vorlesungsgestaltung und die Relevanz der Veranstaltung sind hoch signifikant miteinander korreliert.

5. Diskussion

5.1 Teilnahme am ARS

Der Einsatz von ARS im Rahmen von Vorlesungen wird sowohl von Studierenden als auch von Lehrenden positiv bewertet [5], [6], [7], [8], [9]. In unserer Studie kann keine Aussage zur Akzeptanz des Systems getroffen werden, da diese Frage nicht gestellt wurde. Das mündliche Feedback der Studierenden und der Lehrenden war durchweg positiv. Das Verhältnis von per ARS Evaluierenden zu allen anwesenden Studierenden wurde nicht erfasst. Es wurde jedoch darauf geachtet, dass alle Studierenden ein ARS erhielten und sie wurden in der Lehrveranstaltung mehrfach aufgefordert, dieses einzusetzen. Darüber hinaus wurde das ARS als didaktisches Mittel während der Vorlesung eingesetzt.

Für nicht-anwesenheitspflichtige Lehrveranstaltungen ist eine Abwesenheitsrate von 18,5-70% beschrieben [14], [15], [16]. Wenn man die abgegebenen Stimmen als Zahl der tatsächlich anwesenden Studierende annimmt, dann liegt die mittlere Abwesenheitsrate (mit im Mittel 22 Stimmabgaben pro Frage) im unteren Bereich der in den Studien berichteten Raten. Bei erneuter Durchführung ist das gleichzeitige Erfassen der Anzahl der anwesenden Studierenden zur besseren Einschätzung der Ergebnisse erforderlich.

5.2 Vorwissen, Gestaltung & Relevanz

Der Auftritt und Vortragsstil der Lehrperson kann die Gesamtbewertung einer Veranstaltung, wenn auch nur in geringem Grad, positiv oder negativ beeinflussen [13], [17], [18]. Unsere Daten zeigen ebenfalls, dass vor allem die Vorlesungsgestaltung in einem starken Zusammenhang mit der Gesamtbewertung steht - mehr noch als die Relevanz des Themas selbst. Das studentische Vorwissen steht zwar ebenfalls in einem signifikanten Zusammenhang zur Gesamtbewertung, doch dieser Zusammenhang ist in unserer Untersuchung gering. Es ist davon auszugehen, dass von Studierenden als relevant empfundene Vorlesungen besser bewertet werden [19]. Da diese Einschätzung nicht im direkten Zusammenhang mit dem tatsächlichen Geschehen der Veranstaltung steht, ist davon auszugehen, dass es sich um eine Biasvariable handelt [20]. Die Gestaltung einer Vorlesung hängt dagegen hauptsächlich vom Lehrenden ab [21].

5.3 Zeitlicher Verlauf

In den letzten Wochen des Jahres 2011 im Block 2 fand sich eine signifikant schlechtere Bewertung der Vorlesung. In diesem Block wurden zwei Veranstaltungen besonders kritisch evaluiert. Es ist anzunehmen, dass diese beiden Bewertungen maßgeblich für die Beurteilungsdiskrepanz zwischen den Blöcken sind. Durch die ARS-Evaluation während der Vorlesungsphase können kritisch evaluierte Vorlesungen zeitnah identifiziert und Steuerungsmaßnahmen wie z.B. kollegiales Coaching der entsprechenden

Tabelle 1: Mittelwerte und Standardabweichung der Evaluation mittels ARS und Mittelwert der onlinebasierten Endevaluation (Standardabweichung nicht vorhanden).

Frage	Mittelwert	Standardabweichung
1: „Vorbereitung“	3,18	± 1,24
2: „Vorlesungsgestaltung“	2,44	± 1,16
3: „Relevanz“	2,19	± 1,07
4: „Gesamtbewertung“	2,31	± 1,18
Evaluation Semesterende	2,45	--

Tabelle 2: Vergleich der Evaluationsergebnisse von Block 1 (343≤N≤370) und 2 (405≤N≤463). In beiden Blöcken wurden identische inhaltsgleiche Lehrveranstaltungen angeboten.

Frage	Block	Mittelwert	Standardabweichung	Signifikanz p	T-Test T
1: „Vorwissen“	1	3,07	+/- 1,23	,000	-3,53
	2	3,38	+/- 1,24		
2: „Vorlesungsgestaltung“	1	2,08	+/- 0,99	,000	-5,4
	2	2,51	+/- 1,24		
3: „Relevanz“	1	2,01	+/- 0,98	,002	-3,18
	2	2,24	+/- 1,14		
4: „Gesamtbewertung“	1	2,05	+/- 0,97	,000	-5,54
	2	2,49	+/- 1,28		

Lehrenden oder Überarbeitung der entsprechenden Vorlesungspräsentation ergriffen werden. Im Rahmen einer Semesterende-Evaluation ist dies nur begrenzt möglich.

5.4 Gesamtbewertung inkl. Vergleich zur Semester-Endevaluation

Die Pädiatrievorlesung im Wintersemester 2011/2012 am Universitätsklinikum Freiburg wurde in der Semesterende-Evaluation mit der Note 2,45 evaluiert. In der ARS-Evaluation wurde die Vorlesung mit 2,31 bewertet. Diese Evaluationsergebnisse unterscheiden sich statistisch signifikant voneinander. Gegebenenfalls ist der inhärente Effekt zur positiveren Bewertung bei ARS-Verwendung hierfür mit ausschlaggebend [9], [10], [11], [12].

5.5 Feedback an die Lehrenden

Nach der Evaluation wurden die Evaluationsergebnisse in anonymisierter Form allen Lehrenden präsentiert. Die Lehrenden konnten anhand, nur ihnen bekannter, zugewiesener Nummern die Ergebnisse ihrer Lehrveranstaltung identifizieren. Anschließend wurden die Vorlesungen einem internen Review unterzogen und bei Bedarf überarbeitet. Allen Lehrenden wurde ein Training zur Vorlesungsgestaltung angeboten. Eine erneute Evaluation nach diesen Änderungen steht noch aus, erste Rückmeldungen der Lehrenden und Studierenden sind positiv. Insgesamt waren die Rückmeldungen der Lehrenden über die rasche Verfügbarkeit der Evaluationsergebnisse sehr positiv. Das ARS wurde als geeignetes Evaluationsmedium bewertet.

Tabelle 3: Ergebnisse der Korrelationsanalyse (nach Pearson)
Zahl der Antworten pro Vergleich : 541-810.

Korrelations-Koeffizienten			
	2: „Vorlesungs- gestaltung“	3: „Relevanz“	4: „Gesamt- bewertung“
1: „Vorwissen“	,180**	,147**	,103*
2: „Vorlesungs- gestaltung“		,631**	,836**
3: „Relevanz“			,612**

**: $p = .01$ (2-seitig); *: $p = .05$ (2-seitig)

6. Zusammenfassung

Eine Vorlesungsevaluation mittels ARS ist mit wenig Aufwand während oder unmittelbar nach einer Lehrveranstaltung durchführbar. Darüber hinaus kann das ARS auch zur interaktiven Vorlesungsgestaltung eingesetzt werden [5]. Ein bedeutender Vorteil des Einsatzes von ARS für Lehrveranstaltungs-Evaluationen sind vor allem die innerhalb von Sekunden verfügbaren lehrpersonenbezogenen Ergebnisse. In unserer Untersuchung konnte aufgezeigt werden, dass durch den Einsatz eines ARS als Evaluationsmedium während oder nach einer Lehrveranstaltung, die Variablen „studentisches Vorwissen“, „Vorlesungsgestaltung“ und „Relevanz des Themas“ in signifikantem Zusammenhang mit der Gesamtbewertung einer Lehrveranstaltung stehen.

Einzelne kritisch bewertete Veranstaltungen können zu einer insgesamt schlechteren Evaluation führen. Dies fällt in der ARS-Evaluation sofort auf, so dass hier zeitnah Steuerungsmaßnahmen wie z.B kollegiales Coaching der entsprechenden Lehrenden ergriffen werden können. Durch den alleinigen Einsatz einer Semester-Endevaluation ist dies nicht möglich.

Eine Limitation unserer Untersuchung ist dass die Anzahl der anwesenden Studierenden im Vergleich mit den an der Evaluation per ARS teilnehmenden Studierenden nicht erfasst wurde.

Aufgrund der oben genannten Vorteile sollte die ARS-Evaluation nach weiterer Standardisierung und Ausarbeitung weiteren Fächern zur Verfügung gestellt werden.

Danksagung

Wir danken allen Lehrenden des Zentrums für Kinder- und Jugendmedizin Freiburg und allen teilnehmenden Studierenden.

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

Literatur

1. Kuhnigk O, Weidtmann K, Anders S, Huneke B, Santer R, Harendza S. Lectures based on cardinal symptoms in undergraduate medicine - effects of evaluation-based interventions on teaching large groups. *GMS Z Med Ausbild.* 2011;28(1):Doc15. DOI: 10.3205/zma000727
2. Diehl JM. Normierung zweier Fragebögen zur studentischen Beurteilung von Vorlesungen und Seminaren. *Psychol Erz Unterr.* 2003;50:15.
3. Dresele M, Rindermann H, Tinsner K. Consulting of teachers on the basis of course evaluation of students. 1 ed. Lengerich: Pabst Science Publishers; 2007.
4. Rindermann H. Quality of instruction improved by evaluation and consultation of instructors. *Int J Acad Develop.* 2007;12(2):13. DOI: 10.1080/13601440701604849
5. Caldwell JE. Clickers in the large classroom: current research and best-practice tips. *CBE Life Sci Educ.* 2007;6(1):9-20. DOI: 10.1187/cbe.06-12-0205
6. Rahman A, Jacker-Guhr S, Staufenbiel I, Meyer K, Zupanic M, Hahnemann M, Lührs AK, Eberhard J. Anwendung von elaboriertem Feedback und einem Audience-Response-System in der zahnmedizinischen Ausbildung. *GMS Z Med Ausbild.* 2013;30(3):Doc35. DOI: 10.3205/zma000878
7. Banks DA. Audience Response Systems in Higher Education: Applications and Cases. Hershey: Idea Group Pub; 2006. DOI: 10.4018/978-1-59140-947-2
8. Kay RH, LeSage A. Examining the benefits and challenges of using audience response systems: A review of the literature. *Comp Educ.* 2009;53(3):9. DOI: 10.1016/j.compedu.2009.05.001
9. Savage J. Using an Audience Response System for Module Evaluation. In: Cardiff University (Hrsg). *Technology Enhanced Education.* Cardiff: Cardiff University; 2013. S.5.
10. Plischko R. Mediale Gestaltung von Vorlesungen in der medizinischen Ausbildung mit Hilfe eines Audience-Response-Systems. Dissertation. München: Ludwig-Maximilians-Universität München; 2006.

11. Miller RG, Ashar BH, Getz KJ. Evaluation of an audience response system for the continuing education of health professionals. *J Contin Educ Health Prof.* 2003;23(2):109-115. DOI: 10.1002/chp.1340230208
12. Hecht S, Adams WH, Cunningham MA, Lane IF, Howell NE. Student performance and course evaluations before and after use of the Classroom Performance System in a third-year veterinary radiology course. *Vet Radiol Ultrasound.* 2012;54(2):114-121. DOI: 10.1111/vru.12001
13. Marburger DR. Absenteeism and undergraduate exam performance. *J Eco Educ.* 2001;32:99-110. DOI: 10.1080/00220480109595176
14. Friedman P, Rodriguez F, McComb J. Why students do and do not attend class. *Coll Teach.* 2001;49:124-133. DOI: 10.1080/87567555.2001.10844593
15. Moore R. Attendance and performance: How important is it for students to attend class? *J Coll Sci Teach.* 2003;32:367-371.
16. Abrami PC. Educational Seduction. *Rev Educ Res.* 1982;52(3):446-464. DOI: 10.3102/00346543052003446
17. Marsh HW, Ware JE. Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. *J Educ Psychol.* 1982;74:17. DOI: 10.1037/0022-0663.74.1.126
18. Centra JA. Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness. San Francisco: Jossey-Bass Pub; 1993.
19. Kek M, Stow S. What makes students happy? Factors influencing student engagement using student evaluation data. Brisbane: FYE Curriculum Design Symposium; 2009. S.6
20. Rindermann H. Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen. Landau: Empirische Pädagogik; 1996.
21. Marsh HW. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *J Educ Psychol.* 1984;76(5):54. DOI: 10.1037/0022-0663.76.5.707

Korrespondenzadresse:

Prof. Dr. Marcus Krüger
 Universitätsklinikum Freiburg, Zentrum für Kinder- und
 Jugendmedizin, Mathildenstrasse 1, Freiburg,
 Deutschland
 Marcus.Krueger@uniklinik-freiburg.de

Bitte zitieren als

Bode SF, Straub C, Giesler M, Biller S, Forster J, Krüger M. Audience-response systems for evaluation of pediatric lectures – comparison with a classic end-of-term online-based evaluation. GMS Z Med Ausbild. 2015;32(2):Doc18. DOI: 10.3205/zma000960, URN: urn:nbn:de:0183-zma000960

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2015-32/zma000960.shtml>

Eingereicht: 04.08.2014

Überarbeitet: 03.02.2015

Angenommen: 20.03.2015

Veröffentlicht: 13.05.2015

Copyright

©2015 Bode et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.