

SKURT: Quality Improvement System with Comprehensive Weekly Digital Student Group Feedback

David Sinkvist¹, Annette Theodorsson², Torbjörn Ledin³, Elvar Theodorsson^{4*}

¹Division of Community Medicine, Department of Medical and Health Sciences, Faculty of Health Sciences, Linköping University, Primary Health Care in Linköping, Local Health Care Services in Central Östergötland, County Council of Östergötland, Sweden

²Division of Neuroscience, Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, Department of Neurosurgery, Anaesthetics, Operations and Specialty Surgery Center, County Council of Östergötland, Sweden

³Division of Neuroscience, Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, Department of Otorhinolaryngology in Linköping, Anaesthetics, Operations and Specialty Surgery Center, County Council of Östergötland, Sweden

⁴Department of Clinical Chemistry and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

***Corresponding author:** Elvar Theodorsson, Department of Clinical Chemistry and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden. Tel: +460101033295, +46013286720, +460736209471; Fax: +460101033240; E-mail: elvar.theodorsson@liu.se

Citation: Sinkvist D, Theodorsson A, Ledin T, Theodorsson E (2017) SKURT: Quality Improvement System with Comprehensive Weekly Digital Student Group Feedback. Educ Res Appl: ERCA-124. DOI: 10.29011/2575-7032/100024

Received Date: 28 June, 2017; **Accepted Date:** 28 July, 2017; **Published Date:** 03 August, 2017

Abstract

Students' role in evaluation and rating of teachers and education has been extensively researched for nearly a century. Applied worldwide, students' ratings account for the majority of the available data.

We created a new quality improvement system, SKURT, using digital online weekly combined quantitative, ten-graded scale, and qualitative, open-ended free text, group feedback from medical students. Students rated all educational, non-clerkship, items throughout the entire medical program, spanning eleven terms. The rating process is since 2008 an integral part of a medical program at a Swedish university. The results are, after a screening process, semi-publicly available on-demand, for students and faculty, creating a feedback loop enabling continuous improvement of quality.

A thorough literature search of students rating of teaching found no other corresponding weekly group rating system spanning all educational items. Quality improvement systems based on similar principles as SKURT can uncover problem areas that are difficult to find using other rating systems and has the potential to circumvent several biases, risks and shortcomings of traditional rating systems in current use.

Keywords: Medical Education; Online Evaluation; Problem-Based Learning; Quality Improvement; Rating of Teachers; Student Evaluation

Introduction

Common amongst all methods for improving quality in medical education is the importance of knowledge and insight in the functional and structural strengths and shortcomings at all levels of the medical school, from individual lectures through bedside teaching to organizational issues [1-3]. The students' role in eval-

uation and rating of teachers and education has been extensively researched for nearly a century [4]. Applied worldwide it accounts for the majority of rating data [5]. The term "Student Evaluation" is accepted and widely used but, agreeing with Benton, et al. [6], we prefer using the term "Student Rating" or "Student Feedback".

Student rating of their teaching correlates with student achievements [2,6,7] and the improvement of teaching [8]. Properly used, student rating of teaching has been shown to be reliable, stable, generalizable, valid, relatively free from bias, in line with student achievement and in line with ratings by teachers them-

selves, administrators, colleagues and trained observers [6].

Manners of obtaining student ratings in currently available studies are mainly individual [3,9-29] while the group format [7,30-33] are not as thoroughly evaluated. Ratings are likewise almost exclusively anonymous [3,7,9,11-17,20,22, 25-27,29,33-37] and only seldom anonymous [10,27]. Ratings are most often voluntary [3,11-13,15-20,22,25,29,35,38] while some mandatory [3,10,35,38] rating systems have been implemented in an effort to improve response rates. Questions are either validated [3,7,17,20,22,24,29] or non-validated [11-16,18,19,28] and composed of either solely quantitative [3,7,9,12,13,17,22,24,28], combined [3,7,10,11,14-16,18,19,23,29,33,35-37,39,40] or only qualitative questions [27,32].

The ratings are gathered by paper-and-pencil [3,7,9,10,12,15,17,20,23,26-28,35,38,41] or digital [3,11-19-21-26,29,34-38,40,41] methods. Results are made public [3,13-16,18,19,29], internally public for students and/or faculty [11,13,17,21,29] or closed for administrative purposes only [3,10,21,35,38]. Levels of focus of the ratings ranges from individual educational items [3,7,10,25,27] through whole courses/ per teacher [3,9,11-15,17,18,21,22,24,28,29,35,39] to the entire program [3,7] with frequency from weekly/recurrently [10,11,21,25,27,29,36,37] through mid-term [8,42], end-term-/class [8,9,12,13,15,17,20,24,26] to after graduation [3,7,17,29].

The purpose of the present study is to describe a quality improvement system, SKURT, based on digital online weekly combined quantitative, ten-graded scale, and qualitative, open-ended free text, group feedback from medical students. Students rated all educational, non-clerkship, items throughout the entire medical school, spanning eleven terms. Clerkship sessions are practical training in wards, primary care etc. The results were semi-publicly available for students and faculty at a Swedish university. The system was created to guide formative, quality enhancement, and educational decisions [4].

In this paper we describe the philosophy, technical solutions and practical application of SKURT. In the second paper published simultaneously, we describe the data from and consequences of the use of the system during the five-year period 2009 - 2013.

Method

Context

The medical school has a long-established practice of self-directed problem-based learning [43] with web-based scenarios used in tutorial group sessions in computer- and video-projector equipped rooms. Medical students participate in two-hour long mandatory tutorial groups twice a week. The 6-8 participants of each tutorial group stay the same for the duration of a whole se-

mester (20 weeks). Since 2004 the medical program is organized by seven multi subject cross-term theme-groups who decide on educational activities and examine the students [43].

The medical student association has had a strong standing vis a vi the medical program and faculty, forwarding student opinions and participating in dialogs in different educational forums and influencing policies at program and faculty levels. The students association had two, representative assembly elected, board members with mainly educational responsibilities who served as chairmen for the student quality control of the education and were leaders for a workgroup of students focusing on quality improvement. Furthermore, each theme-group had one or two elected student representatives.

Literature Review

We performed a literature review in order to understand the scope of routines, practical solutions applied, benefits and pitfalls of rating of education in the health sciences. The Education Resource Information Center (ERIC), Academic Search Premier (ASP), PubMed and UniSearch (University search engine combining ERIC, PubMed, ASP and more) were used, in early October 2013 and repeated in April 2014, using keywords and subjects from each database's thesaurus and results as presented in Appendix 1. The reference sections of the relevant articles were searched for additional appropriate sources of information. Only articles published in English were included.

Database	Keyword (s)	No of Results
ERIC	DE "Course Evaluation"	4320
	DE "Course Evaluation" AND DE "Focus Groups"	28
	DE "Course Evaluation" AND DE "Groups"	2
	DE "student evaluation of teacher performance"	3079
	DE "student evaluation of teacher performance" AND DE "evaluation methods"	556
	DE "student evaluation of teacher performance" AND TI review	37
	DE "student evaluation of teacher performance" AND DE "Focus Groups"	9
	DE "Online Surveys"	307
	DE "Online Surveys" AND DE "Student Surveys"	47

ASP	DE "GROUP work in education"	2337
	DE "FOCUS groups"	7602
	DE "GROUP work in education" AND DE "FOCUS groups"	2
	DE "TEACHERS -- Rating of"	2189
	DE "TEACHERS -- Rating of" AND DE "FOCUS groups"	0
	DE "TEACHERS -- Rating of" AND TI review	41
	DE "STUDENTS -- Rating of" AND group	463
	DE "COURSE evaluation (Education)"	456
	DE "EDUCATIONAL evaluation" AND online	352
	DE "EDUCATIONAL evaluation" AND TI review	84
	DE "STUDENTS -- Rating of" AND online	142
	DE "TEACHERS -- Rating of" AND online	69
	DE "TEACHERS -- Rating of" AND DE "INTERNET"	1
PubMed	("Online Systems"[Mesh]) AND "Students"[Mesh]	234
	"Students"[Mesh] AND ("Faculty"[Mesh]) AND "Internet"[Mesh]	188
	((("Online Systems"[Mesh]) AND "Students"[Mesh]) AND "Faculty"[Mesh])	53
	"Students"[Mesh] AND ("Faculty"[Mesh]) AND "Focus Groups"[Mesh]	193
	((("Internet"[Mesh]) AND "Students"[Mesh]) AND "Feedback"[Mesh])	56
UniSearch	SU "problem-based learning" AND SU "student evaluation of teachers"	5
	SU "student evaluation of teachers" AND weekly	44
	SU "student evaluation of teachers" AND "tutorial group"	3

Appendix 1: Literature review sorted on database with keywords and number of results.

Ethical Considerations

Dealing with feedback is fraught with ethical dilemmas [44], especially when a component of grading is included. The SKURT feed-

back was intended to focus on form and content of the educational activity, and not on aspects of the personality of the teachers. The students were informed about this and got feedback on the issue when needed. All feedback was screened before publication.

Technical Aspects

- SKURT was created using the server-side scripting language.
- PHP to access a MySQL open source relational database system.
- The server used IIS on a Windows Server 2008 R2 operating system. The server was hosted within the university computer network.

Development

Two medical students created, on behalf of the medical program, the first version of SKURT the summer 2008. SKURT was beta-tested the fall 2008 by students and administrators in the same term as the developers were studying. Adjustments based on continual direct feedback from fellow students and involved administrators led to program-wide implementation the spring 2009. Continuous dialogue led to improvements and new functions including integrated scheduling, teacher e-mailing and individual teacher rating report pages.

Results

The software was named SKURT as a Swedish acronym for "StudentbaseradeKursUtvärdeRingssystemeT" which translates to "Student based course rating system".

Flow of Data

SKURT served as the hub for the data as depicted in Figure 1 and described below. Numbers in parenthesis denotes step number in Figure 1.

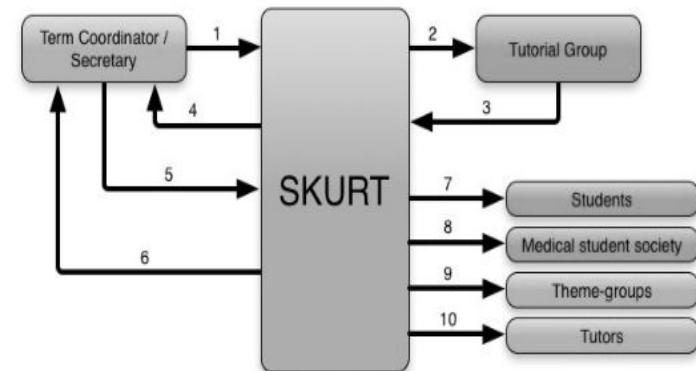


Figure 1: Flow of data in SKURT.

The term coordinator and secretary were the main administrators. The term items were inserted (1) with a smart form auto-filling fields based on previous database entries. Direct duplica-

tion of previous items with subsequent modification of details enabled a simplified input of repeating items. At the end of the last weekly tutorial group session the students logged in to SKURT with a term- and group-specific login and selected a date span for the items to rate (pre-selected as the last seven days). The group was presented with the items available for rating (2). Items not yet rated were preselected but the possibility to edit and supplement previous ratings, labeled with a text snippet, was available. The items were then presented on a single scrollable page one by one with teacher name, item name, type of item, date, time and term.

The items were evaluated using a ten-point scale and a free text comment. Trigger questions aimed to direct the groups comment toward constructive feedback were present above the free-text field. E.g. an information session had trigger questions including “Did you get the information you expected? Did you find anything lacking?” whilst a lecture had the trigger questions “How can this item be improved? How was the time disposition?”. Below the free text form was a ten-point radio button row and a pre-selected “No Points”-button.

Students in each tutorial group alternated in shouldering the responsibility for writing a consensus report for the group based on verbal group discussion of each item. The grading and wording of the written feedback was projected on a screen whilst being written which ensured that each participant could provide hers/his input to the collective feedback. After submitting the ratings (3) a confirmation was shown with a count of number of ratings submitted.

One or two term administrators screened all submitted ratings (4). Each rating was shown on a single row with the students’ comments in a text field. Comments could be revised and administrative comments or feedback could be added in a separate field. Feedback comments and graphic illustration of mean, median, standard deviation and grade distribution were published after screening (5). If comments were changed only the edited version was publicly available and the comment had an italic formatting but not revealing what was revised. The original comment was archived. Administrators could go back and re-screen ratings, enabling screening from several administrators and adding additional administrative comments. Term administrators had both back- and frontend access (6) to the data in SKURT and the only user group able to access unscreened ratings. Each student was provided with a term-specific group login to access ratings for the current, previous and next term (7). The student association board members and the student representatives in the theme-groups had full access to all ratings (8).

Each head of the seven multi-disciplinary cross-term themes groups had full access to all ratings (9). Average grades could be calculated based on date spans allowing weekly or monthly sum-

maries. SKURT was fully searchable on all fields and could easily display trends over time (Figure 2).

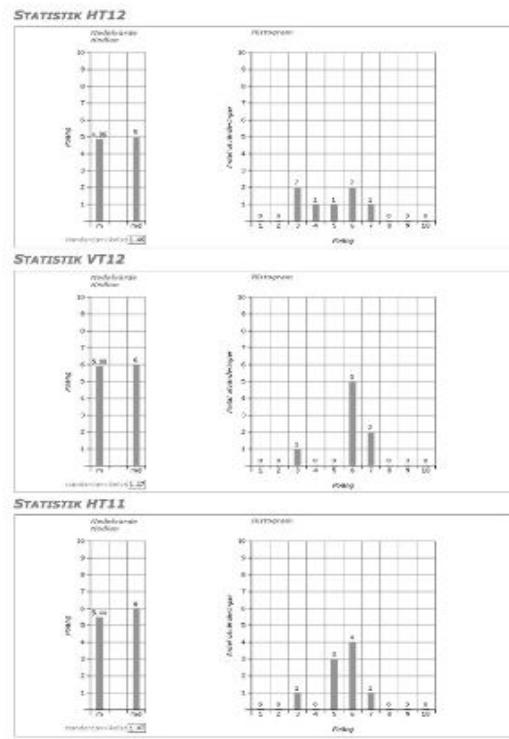


Figure 2: Example of comparison over time for three succeeding semesters with mean, median, standard deviation and distribution of grades.

Each teacher had individual access to all of their associated items ratings including statistical calculations (10). Term administrators could also e-mail individual rating reports to selected or all teachers fulfilling selected criteria. Printing, downloading and exporting rating data were also available.

Scheduling

All items listed in the schedules (Figure 3) (educational items) for all medical students during all 11 semesters were evaluated using SKURT: lectures, tutorials, laborations, seminars and all other items with the exception of clinical tutors in wards and outpatient.

As all educational items to be evaluated needed to be entered into the system, the same information could be used for scheduling. A function for exporting color-coded schedules in Microsoft Word format was implemented (Figure 3). Customization included commenting e.g. “change of premises” or “assignment details”, marking the item as revised, specification of item type and the possibility to create an administrative schedule with otherwise hidden administrative comments.

Vecka 40 - Tema TEST			
Tisdag 1 oktober			
Tid	Ämne	Ansvarig	Lokal
09:15-10:00	[Info] Info från terminsledning		Aulan
10:15-12:00	[Fö] Njursten och akut skrotum Renal Calculus and Acute Scrotum.	Teacher X	Aulan
13:00-14:45	[Fö] Normal bläsfysiologi och miktionsrubbningar Normal bladder physiology and voiding disorders.	Teacher Y	Aulan
15:00-17:00	[BG] Basgruppträff 1		Se TX:s hemsida
Torsdag 3 oktober			
Tid	Ämne	Ansvarig	Lokal
09:15-10:00	[Fö] Syra-Basfall Acid-Base Cases	Teacher Z	Fleming och Haggdahl
13:15-14:00	[Fö] Thoraxtrauma	Teacher P	Victoria
14:15-15:00	[Sem] Seminarium Malignant Hypertension		Victoria
16:00-17:00	[Grupp] Group item Anmälan på TX:s hemsida. Max 16 personer.	Teacher W	Onkologen, O-huset
17:00-18:00	[Grupp] Group item Anmälan på TX:s hemsida. Max 16 personer.	Teacher W	Onkologen, O-huset
Fredag 4 oktober			
Tid	Ämne	Ansvarig	Lokal
08:00-10:00	[BG] Basgruppträff 2		Se TX:s hemsida
10:15-12:00	[Fråg] Question time	Theme group administrators	Nils Holger

Figure 3: Example of schedule.

User Manual

A text-based user manual was available for all users. A combined video and text manual with audio-visual instructions was available for administrators.

Stability and Availability

The choice of software development tools and complete in-house development allowed for extensive flexibility with lean processes including both back- and frontend solutions. The emphases on server-side solutions meant that cross-browser issues were practically non-existent for the benefit of the end-users. No downtime affecting end-users was reported since the launch in 2008.

Discussion

Based on previous experience combined with national and international outlook we created, in 2008, a digital system for combined quantitative and qualitative student group feedback on all educational, non-clerkship, items in an entire medical school comprising thousands of students and faculty and spanning 5.5 years of medical education. The ratings were organized as part of one, of two, weekly tutorial groups. Each tutorial group was expected to join their opinions into a joint feedback in order to prioritize the central tendencies rather than the extremes.

A semi-structured combined quantitative and qualitative approach founded on a single ten-graded scale and an open-ended written feedback field was used to allow for quantitative and qualita-

tive analysis of the complete spectrum of educational items and avoiding possibly irrelevant structured answers in favor of the students' unconfined feedback.

To allow for maximal openness with minimal censorship, the ratings were screened before being published. The direct digital feedback was provided tailored and on-demand for individuals and groups of teachers, students, student organizations, administrators etc. Enabling continuous, first half or mid-term feedback during the term improves quality of teaching as evaluated at end-term [2,6,8,42]. Giving teachers immediate feedback on teaching during the term enables improvements that can be accompanied by new feedback and thus enabling a recurrent open feedback loop with continuous improvement of quality [34,35,37]. Detailed monitoring of the parts facilitates improving the whole whereby continuous rating of each and every educational item ensures quality improvement of the entire program [34].

Today's students are present in both the digital and physical world and teachers and faculty need to be by their side in both. Online digital ratings have obvious administrative, economic and environmental advantages to paper-based solutions [4,6,12,23,26,34,46]. Previous frustrations caused by glitches in IT-solutions and lower response rates have been reasons for preferring paper-based ratings, even without apparent non-response bias and with increasing response rates over time [4,6,34,35,41]. These are and should be glitches of the past and evidence of higher response rates with the online format have been noted [23,35], especially for open-ended questions [6]. Online rating responses are broader and deeper but otherwise consistent with paper ratings regarding mean, correlation and valence [4,6,12,15,23,34,35,40,41,46].

Students not only prefer but also demand convenient digital tools online for making their voices heard and request direct feedback that their opinions and suggestions have impact both digitally and in the real world [1,11,12,26,34,39]. It is crucial not only building a digital rating system but rather include the feedback in a comprehensive quality improvement system where feedback is translated into tangible improvements on all levels. Giving the generators of the data, the students, access to the ratings and improvements ensures a high engagement [2,3,10,11,34,39] while counteracting the need for biased commercial alternatives with accompanying shortcomings [14,15,18]. In concordance with the optimal feedback process the ratings were tailored and available online on-demand for both students, an openness requested [39,47], and teachers shortly after input [34,48,49].

Our medical school relies on problem-based self-directed learning and SKURT as described here is probably best suited for the tutorial group setting, with an online digital learning environment, even though it could be generalized and applied in other pedagogical settings provided student groups have regular meetings with online-access. The program has numerous teachers, most of whom only have a handful of items each semester.

This requires a more item specific rather than course-wide rating. Recurrent ratings augment the item-specific focus with more specific and timely feedback [7,25]. The program also has a fixed curriculum without possibility of selecting individual courses, decreasing the view of students as customers who select high-ranking courses like consumers of education [47,50]. The overwhelming majority of items are voluntary. Both factors decrease the noted bias towards elective courses receiving higher ratings [5].

Tutorial groups meet twice weekly which enhances the habitual use of the system and promotes high response rates, counteracting factors such as forgetting and lack of time noted previously [12,25,26,34,35]. The incorporation of ratings in the mandatory tutorial group setting makes response rates and other factors independent of individual teachers' promotion of rating and potential biasing influencing tactics of teachers [5] should be minimized as the ratings are separate from the items [6].

Students see ratings as a way to improve teaching processes and their outcomes [25]. Ratings serve as both teacher feedback and assessment if learning objectives have been met [39,51]. The usage of SKURT since 2008, integration in the campus culture, involvement of student organizations, improvements based on rating data and student's wide-spread notice of their feedbacks importance ensured high engagement [35,39].

Even though tutorial groups met twice weekly we recommended rating once a week to minimize the risk of "Overuse" resulting in diminished interest or ill-considered feedback [1,11,13,27,39] while at the same time keeping the feedback current and relevant [1,12,34,49] as students also like commenting on a class while still taking it [35]. In comparison with other weekly student ratings [11] SKURT did not consist of repeated ratings, with the same content, in the same class but rather the same questions but on different items, which could counteract the noted overuse and keep motivation high.

The combination of a single quantitative closed and qualitative open-ended rating promotes engagement as students prefer free text comments to scaled ratings, enabling individual feedback to facilitate improvements [39]. The group reaching consensus feedback through discussion facilitate the students' skills in the feedback model [49] and facilitates cooperation with colleagues of other opinions. Skills that are essential in good clinical practice and mentoring [52].

Students have neither expressed nor shown fear of reprisals and are not likely to have expressed exaggerated positive feedback in hope of favors as only group login was registered, examinations were anonymous and not administered by all teaching faculty, all ratings were completed before examinations and examinations were only graded pass or fail; factors which otherwise run the risk of being potential biasing variables [1,2,4-7,15,20,23,47,50,51].

The pre-exam rating timing and close proximity between the teaching item and its rating is in line with effective feedback [1,34,49].

The anonymity of the group ratings setting and online format is of particular importance for students with a correlating increase in honesty of feedback and self-disclosure [6,23,26,34]. No incentives besides improved quality of education was provided for students because of risk for bias [34]. The group-rating used in SKURT could be replaced by individual rating although the group setting should counterbalance individual preferences of learning styles, attendance and other needs with resulting increased potential for feedback being based on quality for the student group as a whole rather than on individual preferences [1,39]. Biasing factors including gender, attractiveness and others [2,5,16,20,23,47] should be compensated in the mixed group setting. The tutorial groups in tutorial rooms with recurrent weekly rating time-slots enables an unprecedented and coveted standardization, of time, place, condition and situation, of digital ratings not previously explored [4].

The initial plan was to enable all students, teachers and faculty to freely browse all published ratings. SKURT and its proposed openness were, and are still not, unequivocally well received. Some faculty and teachers expressed fears that negative ratings would challenge teaching motivation and voices were raised for not making any ratings public for other than main faculty administrators, including making them hidden for the teachers themselves. Fears were expressed that low ratings of teachers would hurt their feelings and make them stop teaching. This might have been the outcome for a minority of teachers but teachers generally seek, welcome and reflect on feedback from students [1,34,36,37,47-49].

All comments in SKURT were seen as opinions rather than stated facts and the guiding questions for the qualitative feedback input box was aimed at promoting effective feedback. The ratings were screened inherently by the group setting and systematically by administrators before being published. The resulting feedback was cleared of conceived prejudiced feedback which could hurt morale and motivation [1]. The screening process was in no way a censorship and significant revision was very rarely used in practice [53]. Unconstructive comments are a risk that has been noted [47] and the students' knowledge of the screening process should minimize this risk. These processes promote constructive feedback [49] and allows for a semi-public quality improvement system with student and faculty access which in turn promotes continuous improvements on all levels [2].

The essence of SKURT was quality improvement on both individual and program level and not to create a ranking system. SKURT or its data was not used or developed into a ranking of the teachers and no function for sorting teachers or items by grade was developed. No monetary award was given teachers for high ranking in SKURT as a high grade should not be a goal in itself [19] but

rather an indicator of educational quality and a guide for professional development. SKURT was not a source for summative decisions as it could lead to e.g. an incentive to water down class content or personnel decisions based on potentially biased data [47]. SKURT could though be used to direct extra pedagogical backing, aiding selected teacher's professional development and identify teachers with excellent teaching skills deserving appreciation [50]. The data was well suited to use in combination with other means of improvement of teaching effectiveness [4,6,48].

Quantitative global ratings correlate well with free text comments [6]. The quantitative data enabled visual trend analysis over time [2] and person and panorama view for administrators. The choice of a ten-point scale was based on the potential flaws of calculating means and statistics on the wide-spread Likert-like scales [28,47]. Discussions if only one quantitative ten-graded scale would be sufficient or if at least two (content and structure) scales would be needed was postponed due to the conceived risk of limiting the generalizability of the system for all activities and as a compromise based on the previously mentioned concerns from a minority of teachers. The open-ended qualitative feedback component helped make sense of the grades. With each tutorial group only giving a single rating per item, the feedback was easily overviewed and grasped.

Using a single grade, a single free text open-ended field, built in regularity and consistency in evaluating all educational items minimizes the risk of noted problems of "Home-Grown" scales [4] and student opinions [34]. The concise format also counteracts students reaching their saturation point for additional ratings [26]. The brief format enables more substantial focus on the qualitative part of the rating compared to other rating systems where the qualitative part is more of a supplement to an extensive Likert-like question and statement list [29].

Future Improvements

Expanding the ratings to include at least two ten-point scales regarding different educational components [6] could refine the quantitative feedback but at the expense of potential overuse of students commitment. An expansion would need a thorough consideration as one or few global ratings can be sufficient for the current purpose [6]. Enabling teachers to respond to the feedback in SKURT could further increase the students' engagement [1] and a function for signaling or reminding the students of missed ratings could further improve the response rate [6].

In Summary

We created a quality improvement system, SKURT, which applies principles that could be applied in all types of curricula. The system uses digital online weekly combined quantitative, ten-graded scale, and qualitative, open-ended free text, group feedback from medical students. Students rate all educational, non-clerkship

items throughout the entire medical school, spanning eleven terms. The rating process is since 2008 integrated in the campus culture and a weekly tutorial group session. The results are, after a screening process, semi-publicly available on-demand, for students and faculty, creating a feedback loop enabling continuous improvement of quality.

The principles applied in SKURT have the potential to circumvent several potential issues noted regarding individual online ratings and can aid in quality improvement on both program and individual level.

Declaration of Interest

The intellectual rights and the copyright to SKURT belong to David Sinkvist.

Acknowledgements

David Sinkvist programmed SKURT and administered all server applications. The three other authors of this manuscript constituted a project group throughout the entire duration of the project, and as heads of the medical program (TL, AT) communicated with teachers (ET, TL, AT), students (DS, ET, TL, AT), institutions (ET) and the Faculty (ET, TL, AT).

References

1. Cleary M, Happell B, Lau ST, Mackey S (2013) Student feedback on teaching: Some issues for consideration for nurse educators. *Int J Nurs Pract* 19 Suppl 1: 62-66.
2. Wright SL, Jenkins-Guarnieri MA (2012) Student evaluations of teaching: combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education* 37: 683-699.
3. Alderman L, Towers S, Bannah S (2012) Student feedback systems in higher education: a focused literature review and environmental scan. *Quality in Higher Education* 18: 261-280.
4. Berk RA (2013) Top five flashpoints in the assessment of teaching effectiveness. *Medical Teacher* 35: 15-26.
5. Pounder JS (2007) Is student evaluation of teaching worthwhile?: An analytical framework for answering the question. *Quality Assurance in Education* 15: 178-191.
6. Benton SL, Cashin WE, Kansas E (2012) Student Ratings of Teaching: A Summary of Research and Literature. IDEA PAPER# 50: 1-22.
7. Richardson JTE (2005) Instruments for Obtaining Student Feedback: A Review of the Literature. *Assessment and Evaluation in Higher Education* 30: 387-415.
8. Cohen PA (1980) Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education* 13: 321-341.
9. Zhao J, Gallant DJ (2012) Student evaluation of instruction in higher education: exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education* 37: 227-235.

10. Youssef LS (2012) Using student reflections in the formative evaluation of instruction: a course-integrated approach. *Reflective Practice* 13: 237-254.
11. Winchester MK, Winchester TM (2012) If you build it will they come?: Exploring the student perspective of weekly student evaluations of teaching. *Assessment & Evaluation in Higher Education* 37: 671-682.
12. Stowell JR, Addison WE, Smith JL (2012) Comparison of online and classroom-based student evaluations of instruction. *Assessment & Evaluation in Higher Education* 37: 465-473.
13. Palmer S (2012) The performance of a student evaluation of teaching system. *Assessment & evaluation in higher education* 37: 975-985.
14. Lewandowski Jr GW, Higgins E, Nardone NN (2012) Just a harmless website?: an experimental examination of RateMyProfessors. com's effect on student evaluations. *Assessment & Evaluation in Higher Education* 37: 987-1002.
15. Legg AM, Wilson JH (2012) RateMyProfessors.com offers biased evaluations. *Assessment & Evaluation in Higher Education* 37: 89-97.
16. Davison E, Price J (2009) How do we rate? An evaluation of online student evaluations. *Assessment & Evaluation in Higher Education* 34: 51-65.
17. Oliver B, Tucker B, Gupta R, Yeo S (2008) e VALUate: an evaluation instrument for measuring students' perceptions of their engagement and learning outcomes. *Assessment & Evaluation in Higher Education* 33: 619-630.
18. Li C, Wang X (2013) The power of eWOM: A re-examination of online student evaluations of their professors. *Computers in Human Behavior* 29: 1350-1357.
19. Palmer S (2012) Student evaluation of teaching: keeping in touch with reality. *Quality in higher education* 18: 297-311.
20. Spooren P (2010) On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching. *Studies in educational evaluation* 36: 121-131.
21. Alderman L, Melanie L (2012) REFRAME: a new approach to evaluation in higher education. *Studies in Learning, Evaluation, Innovation and Development* 9: 33-41.
22. Boerboom TB, Mainhard T, Dolmans DH, Scherpbier AJ, Van Beukelen P, et al. (2012) Evaluating clinical teachers with the Maastricht clinical teaching questionnaire: How much 'teacher' is in student ratings? *Medical teacher* 34: 320-326.
23. Venette S, Sellnow D, McIntyre K (2010) Charting new territory: assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education* 35: 97-111.
24. Bangert AW (2008) The Development and Validation of the Student Evaluation of Online Teaching Effectiveness. *Computers in the Schools* 25: 25-47.
25. Luks AM (2007) An alternative means of obtaining student feedback. *Med Educ* 41: 1108-1109.
26. Donovan J, Mader C, Shinsky J (2007) Online vs. traditional course evaluation formats: Student perceptions. *Journal of Interactive Online Learning* 6: 158-180.
27. Stead DR (2005) A Review of the One-Minute Paper. *Active Learning in Higher Education the Journal of the Institute for Learning and Teaching* 6: 118-131.
28. Huybers T (2013) Student evaluation of teaching: the use of best-worst scaling. *Assessment & Evaluation in Higher Education* 39: 496-513.
29. Tucker BM (2013) Student evaluation to improve the student learning experience: an Australian university case study. *Educational Research and Evaluation* 19: 615-627.
30. Perez JE, Peel JL (1995) Student focus groups: feedback on basic science courses. *Acad Med* 70: 430.
31. Coker J, Tucker J, Estrada C (2013) Nominal group technique: a tool for course evaluation. *Medical Education* 47: 1145.
32. Hamilton DM, Pritchard RE, Welsh CN, Potter GC, Saccucci MS (2002) The effects of using in-class focus groups on student course evaluations. *Journal of Education for Business* 77: 329-333.
33. Fuller HA (1999) The Focus Group as an Effective Tool in College Health Education Evaluation. 1999.
34. Nevo D, McClean R, Nevo S (2010) Harnessing information technology to improve the process of students' evaluations of teaching: An exploration of students' critical success factors of online evaluations. *Journal of Information Systems Education* 21: 99.
35. Anderson HM, Cain J, Bird E (2005) Online Student Course Evaluations: Review of Literature and a Pilot Study. *American Journal of Pharmaceutical Education*: 69: 34-43.
36. Winchester TM, Winchester MK (2014) A longitudinal investigation of the impact of faculty reflective practices on students' evaluations of teaching. *British Journal of Educational Technology* 45: 112-124.
37. Winchester TM, Winchester M (2011) Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development* 16: 119-131.
38. Shah M, Nair CS (2012) The changing nature of teaching and unit evaluations in Australian universities. *Quality Assurance in Education* 20: 274-288.
39. Schiekirk S, Reinhardt D, Heim S, Fabry G, Pukrop T, et al. (2012) Student perceptions of evaluation in undergraduate medical education: A qualitative study from one medical school. *BMC medical education* 12: 45.
40. Champagne MV (2013) Student use of mobile devices in course evaluation: a longitudinal study. *Educational Research & Evaluation* 19: 636-646.
41. Perrett JJ (2011) Exploring graduate and undergraduate course evaluations administered on paper and online: a case study. *Assess Eval Higher Educ* 38: 85-93.
42. Overall JU, Marsh HW (1979) Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. *Journal of Educational Psychology* 71: 856.
43. The students' role in evaluation and rating of teachers and education has been extensively researched for nearly a century[4]. Applied worldwide it accounts for the majority of rating data
44. Norton LS (2012) Action research in teaching and learning: A practical guide to conducting pedagogical research in universities. *Teach Theo Religi* 15: 302-303.
45. Theall M, Franklin J (2001) Looking for bias in all the wrong places: a search for truth or a witch hunt in student ratings of instruction? *New directions for institutional research* 2001: 45-56.

46. van Mook WN, Muijtjens AM, Gorter SL, Zwaveling JH, Schuwirth LW, et al. (2012) Web-assisted assessment of professional behaviour in problem-based learning: more feedback, yet no qualitative improvement? *Adv Health Sci Educ Theory Pract* 17: 81-93.
47. Jones J, Gaffney-Rhys R, Jones E (2014) Handle with care! An exploration of the potential risks associated with the publication and summative usage of student evaluation of teaching (SET) results. *Journal of Further and Higher Education* 38: 37-56.
48. Penny AR, Coe R (2004) Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research* 74: 215-253.
49. Ovando MN (1994) Constructive feedback: A key to successful teaching and learning. *International Journal of Educational Management* 8: 19-22.
50. Gump SE (2007) Student Evaluations of Teaching Effectiveness and the Leniency Hypothesis: A Literature Review. *Educational Research Quarterly* 30: 56-69.
51. Clayson DE (2009) Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education* 31: 16-30.
52. Moss HA, Derman PB, Clement RC (2012) Medical student perspective: working toward specific and actionable clinical clerkship feedback. *Med Teach* 34: 665-667.
53. Sinkvist D, Theodorsson E, Ledin T, Theodorsson A (2014) Five year data and results of continuous quality improvement using SKURT. 2014.