Author's response to reviews

Title: The Quality of Postgraduate Medical Training in Integrative Medicine within the Public Health Care Systems of Germany and Switzerland. The Example of Anthroposophic Hospitals

Authors:

Peter Heusser (peter.heusser@uni-wh.de)
Sabine Eberhard (sabine.eberhard@gmx.de)
Bettina Berger (bettina.berger@uni-wh.de)
Johannes Weinzirl (johannes.weinzirl@uni-wh.de)
Pascale Orlow (pascale@orlow.net)

Version: 3
Date: 28 April 2014

Author's response to reviews: see over
Revision 2

MS: 1219602511546790, BMC Complementary and Alternative Medicine, your Email letter of February 26th, 2014, written by Mr James Prozenko, on behalf of Dr Vincent Chung, concerning our manuscript:

Peter Heusser, Sabine Eberhard, Bettina Berger, Johannes Weinzirl, Pascale Orlow: 
*The Subjectively Perceived Quality of Postgraduate Medical Training in Integrative Medicine within the Public Health Care Systems of Germany and Switzerland. The Example of Anthroposophic Hospitals*

Dear Editor,

Thank you for your Email letter of February 26th and the very helpful comments of the referees and the associate editor. We have completely revised the manuscript and – as communicated with your editorial office on March 17, 2014 – resubmit it herewith before April 30. As requested we answer the comments and questions point by point, whereby all new text elements are emphasized in red:

A) Associate Editor's Comments:

- "Echoing the reviewers’ concern, there are several methodological and ethical issues that should be resolved prior to further evaluation of this submission’s suitability for
publication: The authors aimed to evaluate the “quality” of postgraduate medical training in this paper. It can be expected that the questionnaires were developed to measure quality of education, instead of issues like working situation (as presented in Table 3), degree of integration (table 4) and structural problems (Table 5). In the method section the authors should justify why these are important indicators for educational quality”.

Authors’ answer: we added the following justification: “The questions about the working situation, integration and structural problems were included as important indicators for educational quality because the features they covered were considered to be prerequisites for optimal integrative medical education: a working situation that allows for completing work as well as continuing education during instead of outside contractually agreed working hours, an active integration of conventional and complementary elements through functioning role models provided by trainers as well as in daily practice, and department structures that include regular events explicitly related to integrative medicine such as personal training career supervision, bedside teaching, educational events or study groups as well as the time to take part in these events’”.

• Associate Editor: “Also, how the questionnaire was developed? How the content validity been assessed?” Authors’ answer: to provide this information, we inserted the following passage: “The questionnaire was first developed in 2003 for the annual national assessment of the subjectively perceived quality of postgraduate medical education in all hospitals of Switzerland. Content validity was ascertained in cooperation with the national steering board for postgraduate medical education, consisting of the officers in charge and trainees of the different medical disciplines. After the annual data assessments, item analyses were performed and the optimized questionnaire was submitted to and approved by the steering board”.

• Associate Editor: “It was mentioned that Cronbach’s alpha was calculated but it seems to be unreported in the paper. Authors’ answer: The associate editor may have overlooked that 16 Cronbach’s alpha values are given in figure 1 depicting Global satisfaction with and quality judgment of competency training in conventional (CON) and anthroposophic (AM) aspects of postgraduate medical training in Germany and Switzerland. However, in order to report alpha values completely, we show them in a new table 3 and inserted the following sub-section in the results section: ”Reliability of questionnaire scales. In both countries, in all trainee answers Cronbach’s alpha for the scales of the major dimensions were excellent (>0.90) or good (>0.80) (except for Error Management in Germany) (Table 3); they were also excellent in trainers’ ratings of Global Satisfaction and Clinical Competencies in AM. However, except for Global Satisfaction with CON and Evidence-based Medicine in Switzerland, alpha-values of the trainers were not more than acceptable (>0.70), in some cases even questionable (>0.60) or low (>0.50).”

• Associate Editor: “More information on the trustworthiness of the questionnaire should be reported, preferably with a questionnaire uploaded as an appendix”. Authors’ answer: An upload of the electronic questionnaire is not possible due to copyright reasons: in this case anybody could copy or use the questionnaire, which would be against the law. Concerning trustworthiness see the answer on content validity above.
• **Associate Editor:** “The authors mentioned that ethics approval was not needed. Please provide evidence of exemption from relevant survey research ethics committee.” Likewise: “Thank you for detailing that your study would not have required ethical approval or consent under your local regulations. However, we would ask you to provide us with documentary proof of this exemption.” **Authors’ answer:** The documentary proof of this exemption can be found on the web-page of the Ethics Committee of the Canton of Zürich which is responsible for the ethical approval of all research in medicine carried out by the University and the Federal Institute of Technology in Zürich: „Research projects on human subjects are defined as research projects that affect human subjects physically or psychologically or in whom consequences cannot be excluded. Mere surveys in the sense of opinion-polls are not research projects on humans. Activities in education that are not accompanied by physical or psychological burden do not belong to the definition of Research projects on human subjects” (http://www.vpf.ethz.ch/about/commissions/EK) (Exact translation from German, PH). Thus, in order to make this point more clear, we replaced the last (underlined) part of the following sentence “an ethics approval by an ethics committee and a written informed consent of the participants prior to the assessment had not to be obtained for this study, its focus being sociological and educational, but not medical in character” by “because such a survey for educational quality assessment - which is not accompanied by physical or psychological burden – explicitly does not belong to the definition of research projects on human subjects for which an ethics approval is necessary” and provided it with the reference to the web-site of the ethics committee of the Canton of Zürich.

• **Associate Editor:** “The sub title of “evaluation and statistics” should be renamed as “data analysis”. **Authors’ answer:** this has been done.

• **Associate Editor:** “Appropriate univariate statistical analysis should be performed across group (e.g. German vs. Swiss; number of trainees). Multivariate analysis should also be provided to delineate predictors on “quality” level”. **Authors’ answer:** There is no justification for inferential statistic, because there was no random sampling, but instead a census covering the whole set of trainees and trainers in questions, not only a subset thereof. (Inferential statistics allow to make inferences from the sample to the census or the population. This is not needed in our case.) We can show (almost) exactly whether there was a difference or not; at least assuming that non-response has not created a misleading picture. (In the case of biased responses, inferential statistic would not solve the problem at all, because inferential statistics can only help to take care of random errors, but not systematic errors). Also, the numbers of observations were much too small for a multivariate analysis. It does not make sense to carry out such analyses with observations per cell that are in part considerably below 10. However, the associate editor as well as reviewer Marcus (see below) is right to ask for the calculation of statistical significance to account more precisely for diverse group differences where this was possible (it was only possible for scales whose items were identical). For these reasons we added the following passage in the Data Analysis Section (and modified the contents of the Results and Discussion Sections accordingly): “As our survey was not carried out on random samples or subsets of the targeted populations but consisted of a
census, covering the whole cross-sectional population of trainees and trainers in PGMT in AM hospitals in Germany and Switzerland, interventional statistics such as multivariate analyses were not planned. Also, due to the very small size of the relevant samples (often considerably fewer than ten), multivariate analyses were not possible. Pearson correlation was calculated (two-tailed) to investigate the relationship between specific quality dimensions of PGMT and department size as defined by numbers of trainees per department. Inter-group differences were calculated using the two-sided Wilcoxon matched-pair test for dependent variables and the Mann-Whitney U-test for independent variables (the different distributions deviated from normality). The level of significance was set at $p < 0.05$. In order to calculate the statistical difference between Global Satisfaction scores of trainers and trainees, the scale for trainees had to be adapted (reduction from four to three items) in order to be identical with the scale for trainers (three items). A statistical comparison of Clinical Competency in AM and CON was not possible because of content differences between the different scales, and a statistical comparison of the trainee and trainer scores for Clinical Competency was not possible due to the different direction of the respective questions. To calculate the overall score differences in basic PGMT dimensions between Germany and Switzerland, we included a comparison of departments of comparable size in both countries (seven or less trainees per department, corresponding to seven as being the highest number of trainees in the largest Swiss department), because PGMT dimensions correlated with department size and, in contrast to Germany, Swiss hospitals had only small and medium-sized but no large departments ( predefined as 1-3, 4-10 and 11 or more trainees per department).

- **Associate Editor:** “Baseline characteristics of respondents should be provided in a table”. **Authors’ answer:** In the Results section we added a new small sub-chapter “Baseline characteristics of respondents”, inserted two tables with the baseline characteristics of trainers and trainees, and added the description: "The baseline characteristics of both trainers and trainees differed considerably in both countries. Whereas senior physicians constitute the majority of trainers in Germany, in Switzerland this is true for assistant medical directors and directors (Table 1). Amongst the trainees the mean year of graduation differed by only about three months, but in Germany the completed mean duration of postgraduate medical education was one year longer than in Switzerland and the mean duration of work in their departments eighteen months longer (Table 2). This was due to a small number of relatively older physicians in a resident position for over ten years (N=13, 14.4%), whereas in Switzerland no one exceeded nine years (details not shown). Age was not recorded in order to avoid possible identification and thus violate the anonymity of respondents. In addition, trainees in Switzerland had a more international background (the majority actually coming from Germany), a higher percentage of females and of full-time employees, and most (71%) intended to specialize in general or family medicine, whereas their colleagues in Germany had broader aims for different specialties, the largest portion (24%) for internal medicine.” Also, we have modified the Discussion section accordingly.
• **Associate Editor:** “Design of the tables must be improved such that it is easy to read and understand without having to read all the description below the caption” **Authors’ answer:** We have improved all the tables accordingly (see tables).

• **Associate Editor:** “Please pay special attention to the reviewers’ comments as well as the above for further improvement of the submission,” **Authors’ answer:** The reviewers’ comments are answered below, point by point; those above have all been answered above.

• **Associate Editor:** “Similarly, we would ask you to remove the reference to the declaration of Helsinki from your manuscript. This is because ethics approval is a prerequisite for adherence to this declaration.” **Authors’ answer:** done.

**B) Comments Reviewer 1, Ursula Wolf**

**Major comments**

• **U. Wolf:** “Throughout the entire manuscript please be careful with the terms “integrate” and “integrative medicine” because integrating means one item is integrated into another item - one cannot integrate x and y. Should the authors wish to use “and” they should consider using other terms such as combine”. **Authors’ answer:** done.

• **U. Wolf:** “Limitation of the study need to be discussed (more carefully”): **Authors’ answer:** done.

• **U. Wolf:** “References: One fourth of the references are self-citation which is an amount quite high. Please consider to reduce the number of self-citations”. **Authors’ answer:** We do not think that we can be reproached for citing own necessary literature related to basic and earlier work in this very field. On the contrary: scientifically we feel obliged to cite all of our references that are necessary to report and justify the work on which this paper is based. We would lose this information if we reduced self-citation. Also, we do not think that one fourth is too much. However, we added and updated literature from other sources as well.

• **U. Wolf:** “Although some of the authors are quite fluent in English some parts of the text appear as a word by word translation”. **Authors’ answer:** The manuscript has been checked for English by a professional native speaker and corrected accordingly.

**Abstract**

• **U. Wolf:** “Please differentiate between CAM and AM”. **Authors’ answer:** Done. We inserted a precision clarifying the relation between CAM and AM: “..anthroposophic medicine (AM), a western form of CAM based on CON, offering an individualized holistic IM approach”.

• **U. Wolf:** “The authors refer to evidence based CAM. Is there also such thing as evidence based anthroposophic medicine?” **Authors’ answer:** In the abstract we do not specifically deal with the topic of evidence based anthroposophic medicine (nor is there a need to do that, but more generally with “evidence-based Complementary and Alternative Medicine (CAM)”.


• U. Wolf: “SW:…better teaching culture in AM. Why should that be and what are the indicators?” Authors’ answer: For clarification, “teaching culture in AM” was replaced by “structured formats for supervision and teaching in AM”

• U. Wolf: “GE heavier workload. Why should that be and what are the indicators?” Authors’ answer: Due to the new statistical analyses and their results, this topic is not discussed in the abstract any more. In the main text it is clearly defined: “…heavier workload indicated by less ability to complete work and continuing education within contractually agreed working hours”.

Page 4 Background:
• U. Wolf: “Integration means “one into the other”, see also above. Authors’ answer: We changed “integration of CON and CAM” into “integration of CAM into CON”.

Additionally, care has been taken to change all corresponding wording in the manuscript accordingly.

• U. Wolf: “Most medical schools provide electives…. Be more precise, where, which school, mandatory and elective classes ….little coordination or evaluation. Careful, as conditions may have changed since data generation of ref 13, 14”. Authors’ answer: We have no place and there is no need to go into details (this sentence refers pre- and not to postgraduate medical education), but we updated the references till 2014 and rephrased the sentence: “A majority of those responsible for representing medical schools in Germany, Switzerland and Austria as well as their medical students favor the integration of CAM into the medical system, but only a limited number of courses on CAM have been introduced and evaluated in undergraduate medical education”.

Page 5:
• U. Wolf: “Please give a rational why on ward PGMT should be the most effective form for training and AM compared to e.g. practices and office”. Authors’ answer: This justification was already given, but maybe not convincing enough. We now extend our justification to the sentence: “Indeed, PGMT in an on-ward form can be expected to be the richest and most effective form of training in IM, because it usually provides more systematic, interdisciplinary and scientific learning formats in theory and practice, combined with responsible patient work for more challenging patients, under direct supervision of a more diversified array of experienced trainers in comparison to medical practices. Additionally, AM hospitals usually have a more complete set of specific pharmacological and non-pharmacological AM treatments than practices can offer”.

• U. Wolf: “References 23+24: Both references refer to Switzerland but not to Germany”. Authors’ answer: German reference added

• U. Wolf: “…in a completely integrated way. See above”. Authors’ answer: changed to: “completely integrative way”

Page 6:
• U. Wolf: “Why is it necessary to split and distribute the results in three papers?”. Authors’ answer: We improved the justification for this. The passage now reads: “In addition, in order to provide a differentiated basis for possible improvements to PGMT in AM hospitals, we performed detailed quantitative and qualitative analyses of specific problems in IM training in PGMT as well as problem-solving options from the
viewpoints of trainers and trainees. For reasons of space these results have to be
published in two additional separate papers.”

Method
- U. Wolf: “Please explain or comment on the “age” of the data July to December 2010
and how that relates to the fact that there are meanwhile more trainers and trainees and
also to at least in part other conditions in Germany and Switzerland”. Authors’ answer:
We inserted the following explanation for the age of the data und the changing
conditions in the Methods section (and adapted the discussion Section accordingly):
“The survey was carried out from July - December 2010, and the first data analyses
performed in spring 2011. However, due to a lack of resources and especially of
personnel (pregnancy and motherhood in one case and change of institutions in
another), data processing and the preparation of the manuscript were delayed, and
submission was only possible at the end of 2013. Nevertheless our results still describe
the first and only evaluation of its kind and have remained highly valid for PGMT in
AM.”

Results
- U. Wolf: “What do the authors mean by “more and more frequent”? Authors’ answer:
“a higher rate and a higher frequency of”. We changed the wording accordingly.
- U. Wolf: “More time resources, consider wording2. Authors’ answer: We rephrased:
“have more time to take part”

Discussion
- U. Wolf: “Delete “anthroposophic medicine” and leave only AM without the brackets
because this term has already been introduced”. Authors’ answer: done
- U. Wolf: “Delete reference 28 because no reference is needed here and it is already
explained and referenced before”. Authors’ answer: done
- U. Wolf: “Selection and detection bias instead of selection bias”. Authors’ answer:
done.
- U. Wolf: “An asset of our study ..... Please split this sentence in two sentences”.
Authors’ answer: done.
- U. Wolf: “with complete hospital coverage.. Consider to be more careful in
emphasizing it”. Authors’ answer: rephrased to: “regular hospital coverage”.
- U. Wolf: “Moreover, it is not completely true as e.g. arts therapies, eurythmic therapies
and also interventions provided by nurses are not covered by DRG”. Authors’ answer:
In the process of shortening parts of the discussion this passage has been eliminated.
- U. Wolf: “Generally, among trainees as well as trainers…… I do not understand this
sentence”. Authors’ answer: This passage has been eliminated.

Page 12
- U. Wolf: “First § belongs to the results section. Adapt manuscript accordingly”.
Authors’ answer: done.
- U. Wolf: “Please give number of “longer working hours”” Authors’ answer: this is not
necessary, as this is only a side reference and a more detailed report would expand the
discussion which we are supposed to shorten according to reviewer 2.
- U. Wolf: “Please set the “lower wages than their Swiss colleagues” in perspective to the living costs in the respective countries”. **Authors’ answer:** This passage has been eliminated in the process of shortening part of the discussion.
- U. Wolf: “Reference 34 dates to 2006, are the numbers still valid in 2014?” **Authors’ answer:** The reference has to be related to 2010, not 2014. Also, we didn’t find a newer comparative study.
- U. Wolf: “Reference 35: Please consider the higher living costs in Switzerland”. **Authors’ answer:** This passage has been eliminated in the process of shortening part of the discussion.

Page 13:
- U. Wolf: “One plausible explanation … and thus reduced workload. Please check the logic in this sentence. That does not necessarily correlate. Please change the sentence or provide proof”. **Authors’ answer:** We erased “and thus a reduced workload”.

Page 14
- U. Wolf: “…and other such as anesthesiology… I do not agree with this statement after the semicolon: AM can very well play an important role in anesthesiology and surgery, for example using pre-operative eurythmic therapies and pre-, peri- and post-operative internally, externally applied medication and art therapies”. **Authors’ answer:** Referee is right, we erased the statement.
- U. Wolf: “In general, the better scores …., I do not understand the relatively greater importance of CON than AM. This still allows or even calls for improvement….Please clarify this sentence”. **Authors’ answer:** Rephrased to: “Quantitatively, CON elements play a much more important role in PGMT than AM, the elements of which provide but a complementary extension thereof.”

Page 15: Conclusion
- U. Wolf: “Consider the above mentioned items and adapt the conclusion accordingly”. **Authors’ answer:** The Summary and Conclusions section has completely been rewritten.

Minor comments
- U. Wolf: “Always use the same tense, i.e. past tense”. **Authors’ answer:** done

Abstract
- U. Wolf: “Medical schools offer basic CAM courses…. Please be more precise”. **Authors’ answer:** “courses” replaced by “electives”  
- U. Wolf: “Take in to consideration that PGMT is not only possible in hospitals”. **Authors’ answer:** In the abstract we did not state that PGMT is only possible in hospitals. We just said AM hospitals “train IM in PGMT”. The relation between PGMT in hospitals or practices is dealt with in the text, not in the abstract.
• U. Wolf: “Replace organic with vital”. **Authors’ answer:** According to the native speaker who corrected the English of our text “living”, not “vital”, is the appropriate word. We changed accordingly.

• U. Wolf: “Reference 22: Better provide an English reference, i.e. 20”. **Authors’ answer:** We use both references. 22 is important because it is – albeit available only in German – an extensive Volume on internal medicine from the viewpoint of anthroposophic medicine of 2012, whereas 24 refers to the first book on AM of 1925!

• U. Wolf: “…”AM aspects in a completely integrated way. What is meant by this expression?” **Authors’ answer:** clarified that by adding: “in a completely integrative way, i.e. conceptually as well as practically”.

Page 6

• U. Wolf: “Translate ETH: Federal Institute of Technology”. **Authors’ answer:** As this Institute is repeatedly mentioned in the following text, we preferred to first bring its whole name in English and German and then use the also internationally well-known German abbreviation, thus: “the Swiss Federal Institute of Technology Zürich” *(Eidgenössische Technische Hochschule Zürich, ETHZ)*. Afterwards, we only used the abbreviation ETHZ in order to avoid the constant repetition of a lengthy full name.

Page 7

• U. Wolf: “Consider continued instead of continual education”. **Authors’ answer:** No, the official term is “continuing education”.

• U. Wolf: “Demographic data or personal data instead of personal characteristics”. **Authors’ answer:** Done.

• U. Wolf: “What do the authors mean by “complete survey”?”. **Authors’ answer:** We changed to “full survey” and explained: “full survey, i.e. among all training and trainee physicians in all 11 German and 4 Swiss anthroposophic hospitals”. In other words: we investigated not just a sample of the trainer and trainee population in question, but the whole population.

Page 8: Results

• U. Wolf: “Always use past tense”. **Authors’ answer:** Done.

Page 9:

• U. Wolf: “Very small instead of extremely small”. **Authors’ answer:** Done.

Page 10:

• U. Wolf: “Please consider replacing hardest”. **Authors’ answer:** “hardest” replaced by “most difficult”.

• U. Wolf: “Again continued education”. **Authors’ answer:** No: the official term is “continuing education”.

• U. Wolf: “Again use past tense”. **Authors’ answer:** Done.

Page 11

• U. Wolf: “Study circle consider word”. **Authors’ answer:** rephrased as “study or working groups”
• U. Wolf: “AM instead of anthroposophic”. **Authors’ answer:** Done

• U. Wolf: “What do the authors mean by “more and more frequent”?”. **Authors’ answer:** “a higher rate and a higher frequency of”. We changed the wording accordingly.

• U. Wolf: “More time resources, consider wording”. **Authors’ answer:** We rephrased: “have more time to take part”

---

**Page 12**

• U. Wolf: “Work situation instead of working situation”. **Authors’ answer:** We think we can leave it that way. It is analogous to “working condition”, which is also a correct term

---

**Page 13**

• U. Wolf: “Capable humanistic role models and trainers, check wording”. **Authors’ answer:** Eliminated

• U. Wolf: “About the existence of trainers check wording”. **Authors’ answer:** Eliminated

• U. Wolf: “Study circle, check wording”. **Authors’ answer:** replaced by “study or working groups”

• U. Wolf: “Time resources to take part in them, check wording”. **Authors’ answer:** “resources” has been erased

• U. Wolf: “But gynecology and obstetrics .. consider replacing disciplines with departments and also check the grammar in this sentence, “disciplines were too small”.” **Authors’ answer:** Here the text has been shortened and the passage erased.

---

**Page 14**

• U. Wolf: “Worked out theoretical background, check wording”. **Authors’ answer:** Eliminated

---

**Tables:**

• U. Wolf: The layout of the tables could be improved by inserting a line between the different questions. **Authors’ answer:** We do not understand this comment: all the questions/information in all the tables are separated by lines. However, all tables were improved by including all information necessary for understanding.

---

**References**

• U. Wolf: Please give in brackets an English translation of the German references. **Authors’ answer:** This has been done

• U. Wolf: References in references referring to websites, please include when they were last accessed. **Authors’ answer:** done

---

**Quality of written English:**

• U. Wolf: Needs some language corrections before being published. **Authors’ answer:** See above: the complete manuscript was checked for language problems by a native speaker and improved accordingly.
C) Comments Reviewer 2: Donald M Marcus

Major complete revisions.

- **D. Marcus:** “1. The aim of the study is to evaluate the perceived quality of the integration of anthroposophic medicine into postgraduate medical training in German and Swiss hospitals. The authors do not explain what they mean by “quality”. The data concern the satisfaction of trainers and trainees with their learning environment and working conditions. That is one dimension of quality, but a more important issue is: what competencies are achieved, i.e., what have the trainees learned? In American postgraduate medical training, residents take examinations every year to test their knowledge of the discipline and to identify areas that need improvement. A more accurate title for the paper would be satisfaction with the integrative medicine training, rather than the quality of the training”. **Authors’ answer:** The referee has probably misunderstood the nature of the evaluation. We do not define the quality of postgraduate medical training (PGMT) by the “satisfaction of trainers and trainees with their learning environment and working conditions”, but by a valid and reliable assessment of the following main dimensions as judged by trainees and trainers, dimensions which are officially acknowledged as the elements constitute the overall concept of “quality” of postgraduate medical education in the yearly national assessments of Switzerland and Germany. Quantitatively, the most important dimension refers to clinical competencies: *Clinical Competencies (28 questions), Global Satisfaction with residency training (4 questions), Learning Culture (7), Leadership Culture (6), Error Management (4), Culture of Decision Making (4), Department Culture (4), and Evidence-Based Medicine (4).* The additional questions about the learning environment and the working conditions serve for the identification of factors which can influence the quality of the training. However, this reflects the subjectively perceived quality of PGMT, its most important part is the measurement of the personal judgment of acquired competencies and does not, as the referee rightly says, measure these competencies directly by examinations. In order to make this clearer we change the title to “The subjectively perceived quality of Postgraduate Medical Training...”, and adapted the text accordingly.

- **D. Marcus:** 2. Which of the differences between groups in Fig. 1 and the tables are statistically significant? The meaning if the authors’ statement on p. 7, “As we conducted a complete survey, differences were not calculated statistically”, is unclear. As noted below, the tables lack important information. Assuming that the numbers in the tables represent means and standard deviations, many of the standard deviations are as large as the differences between groups, which suggests that the differences are not statistically significant. The discussion of the reasons for the differences between Swiss and German programs, and the influence of department sizes, is not informative without knowing whether the differences are significant. **Authors’ answer:** In accordance with the referee’s comments we computed anew statistical differences (see Data Analysis section), and we re-wrote the Results, Discussion, Summary and Conclusions sections accordingly (see these sections). For this reason, the whole paper is now greatly improved.

- **Minor essential revisions**
D. Marcus: 1. Figure 1 and the tables are difficult to understand because they lack basic information. The 6-point rating scale is explained for the first time in table 3, which limits comprehension of the figure and tables 1 and 2. Do the numbers in the tables represent means and standard deviations, and do the numbers in parentheses represent the number of questionnaires analyzed? **Authors’ answer:** The referee’s comment “The 6-point rating scale is explained for the first time in table 3” is not justified. In the Data Analyses section (former Evaluation and Statistics section) we had already stated: “The scale format was adjusted to the national school grade systems, i.e. 1 for best and 6 for worst score in Germany, and the inverse in Switzerland. Swiss scales were recoded to the German system for calculation purposes.” Furthermore: all the legends of figure 1 and the tables 1, 2, 3 and 4 contain this information about the 6-point scale and its answer format (see legends). However, for better visibility within the Methods section, we have now placed this information into the sub-section Questionnaire and re-phrased it as follows: “The main answer format for the questions consisted of a 6-point scale corresponding to the national school grade systems, i.e. 1 for best and 6 for worst scores in Germany, and the inverse in Switzerland. Swiss scales were recoded to the German system for calculation purposes.” The referee’s further question - “Do the numbers in the tables represent means and standard deviations, and do the numbers in parentheses represent the number of questionnaires analyzed?” - had already been answered before: in the legends for tables 1 and 2 we had formulated: “Indicated are means, standard deviations (±) and valid numbers of trainees (brackets),” and similar information is provided in the legend of figure 1. The referees may have looked at the figure and tables without taking their legends into account. Nevertheless, the referee is right to demand more detailed information in the tables, and not only in the legends. To improve the clarity of the tables, we have now included the information on means, standard deviations, numbers and p-values in the tables themselves.

D. Marcus: 2. The Discussion is overly long, repetitive and speculative, and should be shortened considerably. **Authors’ answer:** Based on the new calculations of statistical differences and correlations, we have eliminated all speculative elements and based the discussion completely on these results.

**Quality of written English:**

D. Marcus: Needs some language corrections before being published. **Authors’ answer:** See above: the complete manuscript was checked for language problems by a native speaker and improved accordingly.

With these alterations I submit the revision of our manuscript, including the revised figure and tables. In case of further questions please don’t hesitate to contact us again.

Kind regards

Univ.-Prof. Dr. med. Peter Heusser, MME (Unibe)