Reflective Judgment in Medicine –

Developing the Reasoning about Medical Issues Test

BY

GÖTZ FABRY
M.D., Albert-Ludwigs-University, Freiburg, Germany, 1996

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Health Professions Education
in the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Comittee:

Alan Schwartz, Chair and Advisor
Marcia I. Edison
Ara S. Tekian

ACKNOWLEDGMENTS

I would like to thank my thesis committee – Alan Schwartz, Marcia Edison and Ara Tekian – for their patience and their enduring support and assistance during this long journey. I would also like to thank my colleague Marianne Giesler, who helped me with the statistical analyses and has always been a critical friend helping me to improve my work. Furthermore, I would like to thank the DME faculty and staff for creating fantastic opportunities to learn throughout the MHPE program. In particular, I would like to acknowledge Janet Settle for her tremendous support and help with the organizational issues of the MHPE program and thesis process. Finally, I want to thank my students who volunteered to participate in this study and my colleagues in Freiburg and elsewhere with whom I discussed different aspects of my thesis.

GF

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CFA	Confirmatory Factor Analysis
EFA	Exploratory Factor Analysis
FUFM	Freiburg University Faculty of Medicine
GP	General Practitioner
RCI	Reasoning about Current Issues Test
RJ	Reflective Judgment
RJI	Reflective Judgment Interview
RJM	Reflective Judgment Model
RMI	Reasoning about Medical Issues Test

SUMMARY

There is growing awareness that epistemic cognition is a very important conceptual framework – not only but especially in medicine – in order to understand how individuals deal with ill-defined problems. Assessing epistemic cognition quantitatively, however, has proven to be difficult and there is no standard instrument for this purpose to date. Against this background we developed a questionnaire to capture epistemic cognition as outlined in the Reflective Judgment Model. We chose this particular model because it not only explicitly targets how individuals deal with ill-defined problems but also because it is a developmental model delineating how epistemic cognition evolves depending on age, education, experience etc.

Our initial questionnaire comprised 31 items to be rated on a five-point category rating scale. Based on data from a cohort of 313 first and 189 third year medical students we performed confirmatory and exploratory factor analyses. The latter yielded a three-factor solution including 24 of the 31 items that resembled the macrostructure of the Reflective Judgment Model. However, the internal consistency of the scales was rather low (.57 / .65 / .67) indicating that further effort is needed to improve the instrument before it can be used for educational and diagnostic purposes.

1. BACKGROUND

In their everyday work health professionals face two kinds of challenges: well-defined and ill-defined problems (Elstein et al., 1978). While the former can be dealt with on a technical level using propositional knowledge about what the fact of the matter is, the latter are much more complex because “ill-defined problems cannot be defined with a high degree of completeness and cannot be solved with a high degree of certainty” (King and Kitchener, 2004, Eraut 1985). Low back pain for instance, one of the most common medical problems, is a very challenging condition as the correlations between individual symptoms and medical findings is often low, making it difficult to find a specific diagnosis that leads to a respective therapy (Maher et al., 2017). Another example is elevated cholesterol levels in otherwise healthy individuals where there is ongoing controversy whether or not to treat these people (CTT Collaborators, 2012). Some experts argue in favor, some against therapy and both refer to scientific evidence which might leave the individual physician puzzled regarding the decision he or she has to reach with the patient.

To deal with these kinds of problems and to help patients, professionals need to consider a multitude of information from different sources, of different degrees of reliability and weigh these different aspects against the costs and benefits of the potential solutions (Sullivan and Rosin, 2008). Furthermore, in medicine these considerations are usually embedded in a dialogue between physician and patient or even more participants, e.g., relatives and other health professionals, bringing different perspectives, preferences, and values to the process resulting in an even higher degree of complexity (Montgomery, 2006).

Thus, it is very important in medical education to understand how individuals approach ill-defined problems, and building on that, how the competence to deal with these problems can be fostered (Ilgen et al., 2018). However, existing paradigms in medical education do not fully address this issue. The discourse on clinical reasoning for instance, is mainly focused on finding the right diagnosis. It refers mainly to two cognitive strategies: hypothetic-deductive reasoning on the one hand and pattern

recognition on the other hand (Norman, 2005). Regarding medical practice two problems arise: First of all, in some areas of medicine it might be necessary to treat patients without a specific diagnosis (as in many cases of low back pain) (Malterud et al., 2017). Secondly, finding the right diagnosis might not be the biggest medical challenge as "choosing wisely" what to do (which might also embrace to do nothing and "watch and wait") can be much more complex (Cassel and Guest, 2012). While the diagnostic process in many instances is a matter of facts, medical decision making involves much more, i.e., weighing different options against a background of more or less certain scientific evidence, individual experience, patients' preferences etc. (Hunnink et al., 2014).

Another important paradigm that attracts increasing attention with regard to medical practice is reflection. While there seems to be agreement on the importance of reflection for medical education and practice there is a lot of uncertainty regarding conceptual issues, i.e., what reflection actually means (Mann et al., 2009). In a recent critical narrative review on reflection in medical education research, Ng et al. (2015) pleaded for a more thorough conceptual foundation of the concept of reflection and elaborated two major theoretical framings: reflection as epistemology of practice and reflection as critical social inquiry.

For dealing with ill-defined problems, the epistemological framing of reflection is a very important specification as it allows to grasp the core of this challenge more specifically, i.e., evaluating different types and sources of information with regard to their credibility as well as their adequacy to use them for justifying certain claims. Thus, epistemology and epistemic cognition will be elaborated more fully in the next paragraphs.

1.1. Epistemic Cognition

Epistemology as a philosophical discipline is first and foremost concerned with the nature and justification of knowledge (Moser 2002). Building on that, an individual's beliefs about the nature of knowledge ("What is knowledge?") and the process of knowing ("How do we come to know?") are addressed in psychology and education as *personal epistemology* or *epistemic cognition* (Hofer, 2004).

Given the enormous growth and significance of information that individuals have to deal with in their daily as well as in their professional lives, there is a growing interest in better understanding the dimensions, dynamic and development of epistemic cognition. Several conceptual frameworks have been put forward in this regard (Sandoval et al., 2016). Although they differ regarding their breadth and scope they share a common core that can be describe as two dimensions with two respective sub-dimensions as listed in Table I (Hofer and Pintrich 1997).

TABLE I

DIMENSIONS OF EPISTEMIC COGNITION[a]

Dimension	Nature of knowledge "What is knowledge?"		Nature / process of knowing "How do we come to know?"	
Sub-Dimension	certainty of knowledge	simplicity of knowledge	source of knowledge	justification for knowing
Description	fixed or fluid, tentative, evolving	discrete facts vs. interrelated concepts	outside (i.e. authorities) vs. inside (i.e. individual construction)	"What makes a sufficient knowledge claim?": authority, belief, rules, etc.

[a]Hofer & Pintrich 1997

Furthermore, there is also consensus that epistemic cognition develops over time depending on variables such as age, educational level and engagement with specific knowledge-related problems (Hofer, 2001). Again, different conceptual frameworks have been proposed to describe and capture this process that is thought to begin in childhood and continues into adulthood. Educational interventions during school, college or university are especially important to foster this development. Although the

details of the trajectory of epistemic cognition are still controversial, most frameworks assume that at least three typical stages of epistemic cognition can be described that differ with regard to the sources and justification of knowledge (Hofer and Bendixen 2012): 1) A *dualistic* or *absolutist* view at the beginning, where knowledge is perceived as either right or wrong. 2) A stage of *multiplism*, where all perspectives or knowledge claims are regarded as equally valid. 3) A stage of *evaluativism*, where knowledge is perceived as a rule-based individual construction that must be supported by evidence.

1.2. Measuring Epistemic Cognition

Despite the general conceptual agreement about the core dimensions and principal development of epistemic cognition, measuring the construct remains challenging. Several attempts have been made for instance, to reproduce the proposed dimensionality of the construct by means of self-report surveys with Likert-scaled items. However, none of the studies undertaken so far led to unequivocal results (Debacker et al., 2008). Generally, an inconsistency of factors emerged across different studies that seems to be caused by a number of reasons, e.g., flaws in the operationalization of the construct as well as undersized samples. With regard to the operationalization of the construct one challenge is that – compared to other constructs – epistemic cognition is rather abstract or "ego distant". That means, that individuals might never have thought consciously about epistemological issues until they are prompted to fill out a survey. This might lead to artifacts and a high degree of error in the data. Another important issue concerning this matter relates to the question of domain or content specificity of epistemic cognition. There seems to be consensus that some aspects of epistemic cognition are rather general especially at lower or earlier stages of the development or with regard to everyday knowledge (Buehl and Alexander, 2006). Other aspects are rather specific especially at higher or later levels of the development or with regard to more academic knowledge: To substantiate a knowledge claim in a discipline such as mathematics or physics is fundamentally different from the same task in a discipline such as philosophy or history. Thus, especially when it comes to research with adults it might be necessary to use domain-specific instruments which has hardly happened hitherto (Muis et al., 2006).

Results from qualitative studies on the other hand are more satisfying conceptually but lack the possibility for hypothesis testing with regard to individual differences, correlation to other variables such as motivation, learning strategies, etc. (Wood, 1997). However, exactly these questions are important since epistemic cognition is a very promising target construct to address in medical education as it might be a central prerequisite for the competence to deal successfully with ill-defined problems (Eastwood et al., 2017). Thus, to study epistemic cognition in medicine with regard to ill-defined problems it is necessary to develop an instrument that allows to grasp the specific characteristics of these problems as well as typical approaches and strategies to substantiate and justify knowledge claims.

1.3. Dealing with Ill-Defined Problems: Reflective Judgment

Among the well-established conceptual models that have been proposed to capture epistemic cognition, one explicitly targets how individuals deal with ill-defined problems: the Reflective Judgment Model (RJM) (King and Kitchener, 1981). Thus, it seems especially suited for the purpose of this study. Generally speaking, reflective judgment refers to making decisions under uncertainty. The judgment is reflective because finding a solution for the problem at hand is impossible by using formal logic alone. Thus, instead of applying pre-defined rules or algorithms to a finite set of known data, reflective judgment requires first to identify which facts, formulas, and theories are relevant to the problem, then to evaluate beliefs, assumptions, and hypotheses referring to the problem and finally, generating, evaluating and justifying potential solutions against the existing data and other plausible interpretations. This process is basically unlimited and repetitive as new evidence or new hypotheses might make it necessary to revise what has been accomplished so far. However, from a practical point of view, a problem requires a decision on how to deal with it at a certain point in time and thus, reflective judgment can be defined as bringing a (provisional) closure to an ill-defined problem (King and Kitchener 1994, p. 6).

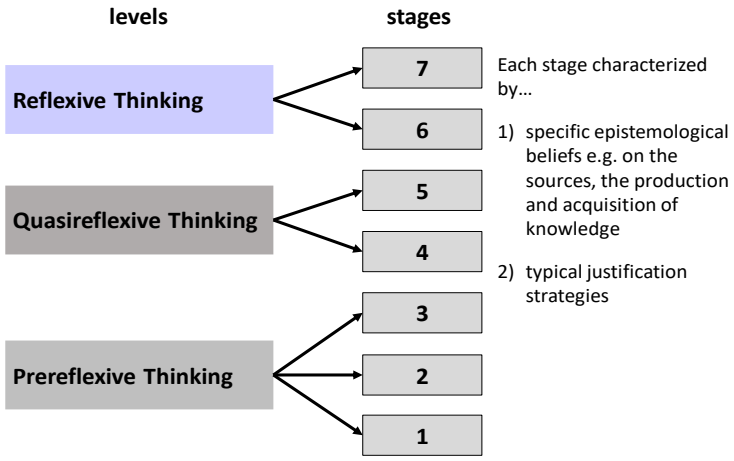
The underlying theory of this model can be traced back to the work of John Dewey (1859 – 1952), who defined reflection as “active, persistent and careful consideration of any belief or supposed form of knowledge in light of the grounds that support it, and the further conclusions to which it tends.” (Dewey

1910, p. 6). Reflection in this sense will be initiated when an "individual recognizes that there is controversy or doubt about a problem that cannot be answered by formal logic alone, and involves careful consideration of one's beliefs in light of supporting evidence" (King and Kitchener, 2004). Thus, reflective thinking occurs only when an individual recognizes that a real problem exists, i.e., when the solution to the problem is not already known. Thus, uncertainty regarding the definition of the problem as well as regarding the accuracy and adequacy of its solution is a constituent element of real, i.e., ill-defined problems. Consequently, a person who does not recognize that a situation is truly problematic (i.e., uncertain or ill-defined) cannot make reflective judgments.

The most important assumption of the RJM is that individuals' assumptions about the nature of knowledge and how they justify beliefs when they are faced with ill-defined problems are logically interrelated. Thus, how an individual explains uncertainty is connected to his or her epistemic cognitions, i.e., the opinion about the sources and certainty of knowledge. Another important assumption of the RJM is that the degree of sophistication characterizing both epistemic cognitions and approaches to justification develops with age and reflected experience in dealing with ill-defined problems. As a developmental model the RJM describes three levels that are further differentiated into seven stages that characterize an individual's capacity to make reflective judgments (Figure 1).

Figure 1

The reflective judgment model[a]



[a]King & Kitchener 2004

Each of these stages is characterized by: (1) a specific view of knowledge, i.e., assumptions about the certainty and potential sources of knowledge and (2) a specific concept of justification, i.e., how people substantiate their claims and whether and how they take alternative views into account (King and Kitchener, 1994). An individual on the *prereflective* level would typically claim something to be a true fact by referring to his or her own perception (level 1: “I’ve seen it, thus, it must be true”) or by pointing to authorities (level 2: ”It was on the news, thus, it must be true”). If this does not lead to a definite answer one must speculate until the answer can be provided (level 3: “If someone can prove it, it is knowledge. Everything else is just guesswork”). At this level, individuals do not recognize that knowledge might be uncertain or that there might be no right or wrong solution for a problem. In contrast, individuals exhibiting *quasireflective* thinking acknowledge that knowledge might be uncertain. However, they believe that this is due to the fact that the respective evidence is not or no longer available (level 4: “Whom shall we ask? We will never know”). They do also acknowledge that there might be more than one perspective on a problem but they claim that this is caused by an idiosyncratic selection of arguments or because knowledge is always bound to a specific context (level 5: “People have different opinions and thus, they deal differently with problems”). Finally, individuals on the level of *reflective* thinking would typically substantiate a claim by arguing that knowledge is an individual

construction following certain rules using evidence that might be flawed by uncertainty (level 6: “It is very difficult to know something for sure. Some sources are more credible than others and eventually you can be sure enough to build your personal opinion”). Thus, weighing the different options one might still come to a conclusion that has the best explanatory power albeit temporarily (level 7: “You can evaluate an argument with regard to the kind of evidence used and the quality of the rationale”).

1.4. Measuring Reflective Judgment

Most commonly reflective judgment has been measured qualitatively by means of the semi-structured Reflective Judgment Interview (RJI) (King and Kitchener, 1994). During the RJI participants are confronted with a number of short vignettes of ill-defined problems and are prompted to answer open-ended questions, e.g., how they would explain that experts hold opposing views on a certain issue. The answers are scored by trained raters using the Reflective Judgment Scoring Rules.

In modified form the RJI has also been used in medical education research. Roex et al. (2009) used ill-defined problems from the RJI as well as newly developed ill-defined medical problems to elicit the epistemic cognitions of trainers and trainees in a GP program in Belgium. However, participants were not interviewed individually but within two focus groups. Statements referring to epistemic cognitions were then scored according to the scoring rules from the RJI. Results show that participants were mainly arguing on the pre- and quasi-reflective level. Furthermore, the exposed level of epistemic cognition differed considerably between different problems suggesting a large degree of content specificity. In a study at one US dental school, Boyd (2008) interviewed undergraduate students three times consecutively over the course of their third year. Instead of using predetermined ill-defined problems as the stimulus material, students were prompted to report and elaborate on critical incidents they had experienced. Again, statements referring to epistemic cognition were coded using the scoring rules of the RJI. Results show that most of the students developed a higher level of reflective judgment over the course of their third year (4.89 to 5.59) which translates to progress from the quasi-reflective (stage 5) to the reflective level (stage 6).

Since the RJI is an instrument using qualitative data, it is not suitable for use with larger groups of individuals. Thus, an objectively scored measure of reflective thinking has also been developed, the Reasoning About Current Issues Test (RCI) (Wood, Kitchener & Jensen 2002). In this test participants are also presented with ill-defined problems. However, instead of using open ended questions, the RCI prompts participants to rate and rank order statements that they think mirror their own views on the issue at hand most closely. Each of the statements, which were derived from participants' responses in prior studies using the Reflective Judgment Interview, reflects one of the stages of the RJM. However, the RCI has hitherto not been used in medical education and publications of studies using this instrument are rare overall. This might at least partly be due to the fact that the instrument itself was never published.

Thus, while the RJI due to its focus on dealing with ill-defined problems seems to be a very promising conceptual model for studying epistemic cognition in medicine and medical education, no instrument exists as yet to pursue quantitative research questions. Against this background the goal of this study is to develop the *Reasoning about Medical Issues Test (RMI)*, which is supposed to be suitable for measuring reflective judgment in medical contexts with larger groups.

2. METHODS

2.1. Development of the Test

As outlined above, the quality of an individual's reflective judgment according to the Reflective Judgment ModeI (RJM) has mainly been determined by scoring the statements made during the Reflective Judgment Interview (RJI) and assigning them to the levels of the RJM. Typically, the statements then reveal the predominant level of that individual's reflective judgment. Building on that, the central idea of the Reasoning about Medical Issues Test (RMI) is to present individuals a number of typical statements that represent each level of the RJM and prompt them to rate these in terms of their agreement. The level of agreement on statements that express the predominant level of the participants' reflective judgment is expected to be higher than the agreement on other statements. Thereby, it should be possible to determine the individual level of reflective judgment by calculating respective scores. Furthermore, if the statements can be assigned to one level of the model with sufficient reliability, it should also be possible to check whether the RJM can be reproduced empirically.

As mentioned in the introduction, evidence prevails that epistemic cognition is rather context specific, especially with regard to academic domains. Thus, the stimulus material used for the RMI is to be related to typical ill-defined problems in medicine. Furthermore, the test items should capture typical epistemic challenges in dealing with ill-defined problems in medicine rather than general statements that run the risk of being too abstract or "ego-distant".

To develop the test-items we choose three very common ill-defined problems that are intensively and repeatedly discussed in the medical literature: 1) Whether or not to treat otherwise healthy individuals that have an elevated cholesterol level with statins (Redberg and Katz, 2012; Blaha et al., 2012); 2) Whether or not "idiopathic low back pain" is an acceptable diagnosis after specific organic causes have been ruled out (Abraham and Killackey-Jones, 2002; Deyo, 2002); 3) Whether or not to recommend regular breast cancer screening for women between the age of 30 and 60 (Broeders et al., 2012). We developed short vignettes (between 150 and 300 words in length) to illustrate these problems and to specify the problem or the respective controversy. Medical students in their first year were then invited

to comment on these vignettes (anonymously, as a voluntary course assignment). They responded to two questions that were derived from the RJI: the first one asked for their personal opinion regarding the controversial issue (e.g. whether or not to treat the patient) and the second one prompted them to explain how they think it is possible that experts in the same field disagree about the respective issue. 20 students commented on the vignettes.

Using the comments of the students as well as the examples provided in the literature on the RJM we phrased 35 prototypical statements that would represent the different levels of the reflective judgment model (see Table II).

TABLE II
REFLECTIVE JUDGMENT MODEL, EXAMPLES OF ITEMS FROM THE TEST DEVELOPMENT

Level	Stage	Example of Statement
Prereflective thinking	1	"I once had to take statins myself and had a horrible experience. Since then I am very restrictive in prescribing these drugs."
	2	"If all experts were honest such controversies wouldn't exist."
	3	"Different opinions exist because we do not know enough yet. Thus, the whole thing is just a matter of belief."
Quasireflective thinking	4	"As long as we don't know the definite answer everybody can bend the evidence to make it fit for his or her own purpose."
	5	"Different positions arise since knowledge always has to be constructed on the basis of theories and data."
Reflective thinking	6	"It is very difficult to know something for sure. Eventually you can be certain enough to form a personal opinion on a specific issue."
	7	"Since definite answers in medicine are rare one must evaluate a certain viewpoint e.g. with regard to how plausible the argumentation is based on current evidence."

To check the content validity of the statements we invited four experts with a background in medical education and pedagogical psychology who were briefed about the theory underlying the RJM, to assign every item to the stage they felt it would refer to. We only kept those items that at least three of the

four experts (75%) had assigned to the same level of the RJM. We used the remaining 31 items for the initial version of the test. Table III shows the resulting number of items for each of the levels and stages of the RJM.

TABLE III
DISTRIBUTION OF ITEMS ACCORDING TO THE EXPERT RATINGS

Level	Stage	Number of items (∑=31)
Prereflective thinking	1	4
	2	3
	3	4
Quasireflective thinking	4	4
	5	8
Reflective thinking	6	5
	7	3

The test consists of an introduction (one page) to explain the background, i.e. discussing ill-defined problems in medicine (see appendix). To avoid prompting of the participants the term “ill-defined problems” is not used, instead, two examples for ill-defined problems – statin therapy and breast cancer screening – are used to illustrate that controversial issues in medicine exist where even experts disagree in their opinion. Then, the explanation of the test instruction follows: Participants are prompted to evaluate the 31 statements with regard to how appropriate they think the statements are for either justifying one's own point of view or explaining the diversity of opinions regarding problems like the two examples. On the next page the items are then to be rated on a five-point category rating scale with the extremes named: “very inappropriate” and “fully appropriate”. In the final section of the test

participants are prompted to choose up to three items that are most similar to their thinking and to rank them according to their importance. We added this element which has also been used in the RCI as a supplementary indicator for the preferred level of reflective judgment.

Finally, the first page of the test contains information on the background of the study, questions for basic sociodemographic data, as well as a coding form that allows for anonymized longitudinal within subject analyses.

2.2. Test Scoring

The RJM is a developmental model suggesting that the epistemic cognitions that become apparent when an individual scrutinizes an ill-defined problem are predominantly situated on one level. However, since the RJM is a "complex stage theory" an individual might also express beliefs or cognitions that relate to other stages as well (Wood 1997). Thus, we developed a scoring procedure that – provided the conceptual model would be reproduced empirically – accounts for these premises. This happens by rescaling the item values and allocating them specific weights that represent the level of reflective judgment they are assigned to conceptually.

Participant use a 5-point category rating scale on each item to indicate their opinion on this statement. Values between 1 and 5 are assigned depending on the answer provided. In a first step subscores for each of the assumed seven scales (Sc) are calculated by averaging the item values within each scale.

$Sc1 = (item\ 1_{Sc1} + Item\ 2_{Sc1} + item\ 3_{Sc1} + item\ 4_{Sc1} + item5_{Sc1}) / 5$ etc.

In a second step an overall score for the RMI is calculated as follows. Because the RJM is a hierarchical model, higher levels represent a more sophisticated and differentiated argument quality and use. Thus, more advanced and experienced students are expected to exhibit higher levels of reflective judgment that should also lead to higher scores in the RMI. To guarantee that, the following transformations are made in order to calculate the overall score of the test: item values are rescaled and weighted to represent the relative importance of their respective judgment level.

1) In a first step the relative importance of each scale is calculated by the following procedure:
   - Item values for each scale are summed up (S_1 = item 1_{Sc1} + Item 2_{Sc1} + item 3_{Sc1} + item 4_{Sc1} + item 5_{Sc1} etc.)
   - After subtracting the scale minimum (Number of items x 1) sums are divided by the maximal scale value (Number of items x 5) minus the scale minimum:

     $S_{1res} = (S_1 – 5 / 25 – 5)$ etc.
   - The rescaled sums (S_{res}) of the scales are divided by the total of the rescaled sums:

     $S_{1rel} = S_{1res} / S_{restotal}$ etc.

   The resulting value (S_{rel}) represents the relative importance associated with each scale.

2) To ensure that the individual test score points to the preferred level of judgment of that person it is necessary to assign specific weights to each scale that represent the level they belong to in the RJM. Thus, the relative weight (S_{rel}) of each scale is multiplied with the rank of each level (represented by that particular scale) in the RJM where level 1 is assigned the value 1 and level 7 is assigned the value 7. The total of the resulting products represents the preferred judgment level (J_{pref}) of the individual:

   $J_{pref} = (S_{1rel} * 1) + (S_{2rel} * 2) + (S_{3rel} * 3) + (S_{4rel} * 4) + (S_{5rel} * 5) + (S_{6rel} * 6) + (S_{7rel} * 7)$

According to that scoring procedure an individual with a test score of Jpref = 4.3 for instance, argues predominantly on the quasireflective level (stage 4).

2.3. Setting and Participants

Medical Students in the first and third academic year at Freiburg University Faculty of Medicine (FUFM) were asked to participate in the study which was conducted during the winter semester of 2016/17. FUFM enrolls approximately 330 students once every year who then pass through a six-year curriculum (two pre-clinical and four clinical years). Students in the first year were asked to participate after they had attended a scheduled lecture. The vast majority of students in this cohort does attend this lecture. Students in the third year were invited after they had participated in an end-of-term assessment. Since

the assessment takes place in three different rooms, the student cohort is split randomly into three groups of equal size (N ≈ 110 respectively). For organizational reasons two of the three groups were invited to participate in the study. It was explained to the students that participation in the study is completely voluntary and that they could opt out anytime if they would wish to do so. The test was provided as a paper-and-pencil version. After finishing the test, participants put their copies in two boxes so that their anonymity was guaranteed.

2.4. Statistics

Because we used the Reflective Judgment Model as the conceptual framework for the development of the test we planned to conduct a confirmatory factor analysis in order to test whether the data fit the hypothesized seven stages or the three levels of the RJM respectively. However, because this was the first attempt to develop items that reflect the RJM we also considered an exploratory factor analysis in case these models would not be confirmed.

SPSS 24 and SPSS Amos 24 were used for the statistical analyses. We tested three different models with confirmatory factor analyses (table 4): 1) the original model, i.e. seven stages nested in three levels; 2) a three level model and 3) a seven stages model. The exploratory factor analysis was performed as a principal axis factoring followed by Varimax rotation. Missing data were excluded list-wise. We considered a factor as interpretable if it is defined by at least three items having loadings of .30 or greater and sharing at least half of their communality with this factor (Fürntratt, 1969). Items meeting these criteria were then assembled into subscales. Internal consistency of the scales was estimated via Cronbach's α.

3. RESULTS

313 first and 189 third year students participated in the study. The average age is 21.1 (SD 3.9) and 22.9 (SD 9.5). In the first year 203 students are female (65%) in the third year 126 (67%). This proportion resembles that of the two student cohorts invited for the study.

3.1. Factor Analyses

Table IV shows the indicators of the confirmatory factor analysis. The values do not indicate a good model fit: the factor loadings are mostly ≤.30 (range .023 – 0.40). Only four items had loadings ≥ .30.

TABLE IV
INDICATORS OF THE CONFIRMATORY FACTOR ANALYSIS

	χ^2	df	p	$\chi^{2/}$df	GFI	AGFI	NFI	TLI	CFI	RMSEA
cut-off Scores			>.05	<2	≥.90	≥.90	≥.90	≥.90	≥.90	≤.06*
Original Model (Seven Stages nested in three levels)	1531.055	413	.000	3.7	.84	.81	.50	.51	.57	.07
Three-level model	1671.505	431	.000	3.9	.79	.759	.38	.41	.45	.08
Seven-stages model	1392.877	413	.000	3.4	.84	.802	.49	.51	.56	.07

(A)GFI = (Adjusted) Goodness of Fit
NFI = Normed Fit Index
TLI = Tucker Lewis Index
CFI = Comparative Fit Index
RMSEA = Root Mean Square Error of Approximation
*if N > 250

The exploratory factor analyses yielded three factors explaining 21% of the variance. Due to factor loadings < .30 six of the 31 items were excluded. One other item was excluded because of double loadings (No. 16). Table V presents the rotated loadings of the items. Reliability and scale indices are shown in Table VI.

TABLE V

RESULTS OF THE EXPLORATORY FACTOR ANALYSIS, FACTOR LOADINGS OF THE ITEMS

No.	EC*	Item	F1	F2	F3	h^2	a^2/h^2 F1	a^2/h^2 F2	a^2/h^2 F3
7	QL	Faced with multifaceted findings from scientific studies you have to arrive at your own conclusions and since people are different their interpretations are different too.	**.36**	.04	.27	.20	**.64**	.01	.36
9	RL	Even experts hold different opinions as they evaluate the available evidence differently. However, some conclusions are more appropriate than others and reflect a more comprehensive synthesis of the available information.	**.44**	.04	.27	.20	**.96**	.01	.02
15	QL	Presumably, we will never agree on such issues as the scientific evidence is constantly evolving and changing.	**.31**	.20	.20	.17	**.56**	.23	.23
17	RL	Since definite answers in medicine are rare one must evaluate a certain viewpoint e.g. with regard to how plausible the argumentation is based on current evidence.	**.45**	.17	.30	.24	**.86**	.14	.01
19	PL	Uncertainty with regard to medical decisions is primarily a matter of insufficient experience and knowledge.	**-.31**	.26	.14	.18	**.53**	.37	.11
20	RL	It is very difficult to know something for sure. Eventually you can be certain enough to form a personal opinion on a specific issue.	**.31**	-.02	.25	.16	**.59**	.00	.39
22	QL	No matter how competent, experienced or knowledgeable you are: uncertainty regarding medical evidence will never completely dissolve.	**.55**	-.02	-.10	.31	**.96**	.00	.03
24	QL	The different viewpoints are caused by different perspectives: A biochemist takes a different perspective than a sociologist.	**.37**	.26	.23	.26	**.53**	.26	.19
25	RL	Knowledge in general is fluctuating and prone to uncertainty.	**.57**	.05	.01	.33	**.98**	.01	.00
29	RL	It's in the nature of medical problems that they don't have a definite answer.	**.45**	.05	.18	.24	**.84**	.01	.14
30	RL	While there is no such thing as a definite truth in medicine you can still assess the quality of a specific proposition or claim e.g. in terms of plausibility, transparency and credibility.	**.47**	-.16	-.15	.27	**.82**	.10	.09
4	QL	I don't trust results from drug studies because they are sponsored by the pharmaceutical industry and thus, the outcomes are biased in favor of the respective manufacturer.	.07	**.62**	-.06	.39	.01	**.98**	.01
5	QL	The major reason for the different opinions is that some experts present the evidence from studies in a biased way in favor of specific interests.	.04	**.57**	-.05	.33	.01	**.97**	.01
6	RL	I trust the official medical guidelines because they are a result of a rigorous and systematic review process from experienced experts.	.07	**-.34**	.02	.12	.04	**.95**	.00
13	PL	If all experts were honest such controversies wouldn't exist.	-.20	**.48**	.17	.30	.13	**.78**	.10
21	QL	Since there is so much evidence on medical issues it is quite arbitrary what one uses to justify a certain position.	-.11	**.35**	.33	.24	.05	**.51**	.44
31	QL	Unfortunately, many experts are not credible because they don't act independently from third party or commercial interests.	.20	**.58**	-.11	.32	.13	**1.05**	.04

28	QL	As long as we don't know the definite answer everybody can bend the evidence to make it fit for his or her own purpose.	-.02	**.36**	.33	.24	.00	**.55**	.46
8	PL	Before I become too confused by controversial findings from studies I prefer to rely on my personal experience with patients.	.04	.14	**.43**	.21	.01	.10	**.90**
10	PL	Different opinions exist because we do not know enough yet. Thus, the whole thing is just a matter of belief.	-.05	.28	**.39**	.24	.01	.34	**.64**
11	PL	On a recent conference Professor X who is a renowned expert in the field advocated statin therapy very much. I will follow this advice.	-.15	-.16	**.37**	.19	.12	.13	**.74**
18	PL	I have personally seen patients dying whose life could have been saved by a timely therapy. Thus, I am in favor of screening tests.	.08	.03	**.44**	.20	.03	.00	**.95**
26	PL	I once had to take statins myself and had a horrible experience. Since then I am very restrictive in prescribing these drugs.	.14	.08	**.40**	.19	.11	.04	**.86**
27	PL	When there seems to be more than one answer I usually adhere to what I learned during my residency in XY which so far has never failed me.	-.01	-.03	**.44**	.19	.00	.00	**1.02**
Excluded items:									
1	PL	I have treated many patients with statins and never witnessed any severe side-effects.	.11	-.06	.24	.08	.16	.04	.73
2	QL	Certainly, other individuals might come to alternative conclusions if they pick the "right" arguments for them.	-.02	.20	.28	.12	.00	.34	.67
3	PL	Such controversies arise because we do not know enough yet. Future evidence will prove who is right.	.01	.08	.08	.01	.01	.67	.59
12	QL	It all depends on the perspective you take because this determines how e.g. the results from different studies are interpreted and used to build an argument.	.29	.23	.25	.20	.41	.26	.32
14	QL	Different positions arise since knowledge always has to be constructed on the basis of theories and data.	.16	.03	.08	.03	.89	.02	.20
16	QL	Medical science is not just based on facts. Personal views and opinions do also matter resulting in different points of view.	.31	.17	.30	.21	.45	.13	.41
23	PL	I rely on medical textbooks very much because they are written by experts in the field.	.00	-.15	.25	.09	.00	.24	.72

N=489; Items are grouped factor-wise. Excluded items are listed at the bottom of the table

h^2 = communality

a^2/h^2 = Proportion of the communality explained by the respective factor (a=loading). Should be >.50

*EC = Expert Categorization i.e. level of the Reflective Judgment Model the Expert had assigned this item to: PL = Prereflective Level / QL = Quasireflective Level / RL = Reflective Level

TABLE VI
RELIABILITY AND SCALE INDICES

	explained variance (%)	Number of Items	Scale Mean	Standard-deviation	Discriminating power (range)	Cronbach's α
Factor 1 ("complex reasoning")	7.64	11	3.84	5.21	.18 - .46	.67
Factor 2 ("skeptic reasoning")	6.82	7	2.54	4.08	.26 - .45	.65
Factor 3 ("experience based reasoning")	6.38	6	2.56	3.63	.17 - .38	.57

The items loading on factor 1 essentially express three major themes: uncertainty as an inherent quality of medical knowledge and decisions (No. 15, 22, 25, 29), reasons for individual differences in the use of arguments and evidence (No. 7, 19, 24) as well as the need and possibility to evaluate the quality of arguments regarding their plausibility and their use of the available evidence (No. 9, 17, 20, 30). Thus, we named factor no. 1 "complex reasoning". In contrast, factor 2 embraces items that mainly express skepticism either regarding the honesty or neutrality of experts (No. 5, 13, 31 and – with a negative loading – No. 6) or the credibility of results from scientific studies (No. 4). In line with that, other items express mistrust in the integrity of the way arguments are used (No. 21, 28). Thus, factor No. 2 was labelled "skeptic reasoning". Finally, items loading on factor 3 mostly relate to the role of personal experience as a guideline for medical decisions (No. 8, 18, 26, 27). The other items bring up experts as credible sources of knowledge (No. 11) and individual idiosyncrasies as a reason for differing opinions (No. 10). Thus, we named factor No. 3 "experience-based reasoning".

3.2. Comparison with Expert Mapping

We also reviewed to which of the stages of the RJM the items of the three factors had been assigned to by the expert panel during the test construction phase (table VII). The majority of the items of factor 1 had been assigned to the reflective level, the items of factor 2 largely to the quasireflective level and all of the items of factor 3 to the prereflective level. It is important to note that two items (No. 19 / Factor 1 and No. 6 / Factor 2) have negative loadings.

TABLE VII

NUMBER OF ITEMS IN EACH OF THE THREE FACTORS WITH REGARD TO THE LEVEL AND STAGES THEY WERE ORIGINALLY ASSIGNED TO BY EXPERT RATING

Level (according to RJM)	**Prereflective**			**Quasireflective**		**Reflective**	
Stage (according to RJM)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Factor 1 ("complex reasoning")			1*		4	4	2
Factor 2 ("skeptic reasoning")		1		3	2		1*
Factor 3 ("experience based reasoning")	3	1	2				

*This item has a negative loading.

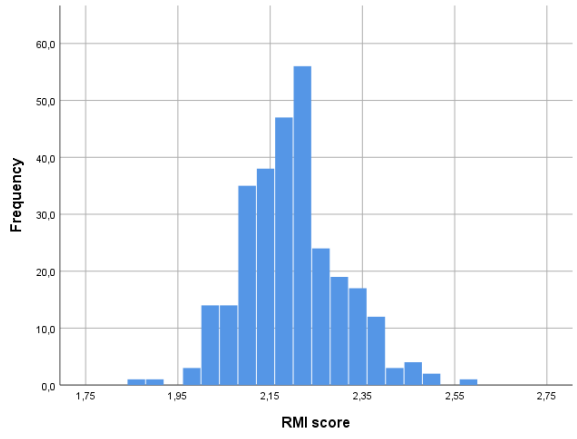
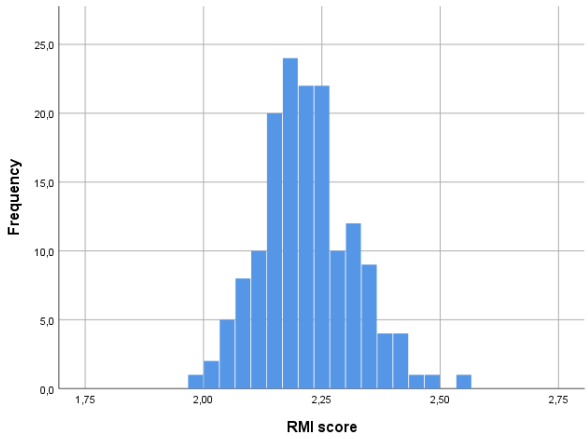
3.3. Test Scoring

Regarding the calculation of the preferred level of judgment we adapted the algorithm depicted in the method section to adjust it to the three factor solution. Thus, we rescaled and transformed the item

values for the three empirically derived factors instead of the seven postulated levels. As a consequence, the maximal value for the preferred level of judgment is 3, indicating complex reasoning, the minimum is 1, indicating experience-based reasoning. Among the first year students the preferred level of judgment is 2.19 (SD 0.14) and among the third year students it is 2.22 (SD 0.10). The distribution of the RMI scores is shown in Figure 2. The difference is statistically significant (independent-samples T-Test: t(447)=-2.445, p=0.015), the effect size (Cohen's d) is 0.19 (95% CI: 0.007 – 0.378).

Figure 2

Distribution of the RMI scores

1st year students (N=293)	3rd year students (N=156)
	

Further analyzes turned out that the preferred level of judgment did not differ with regard to gender, age or prior experience within a health profession.

3.4. Students' Pick of Items Similar to their own Thinking

We also analyzed which statements students picked because they deemed them to be similar to their own thinking. Table 8 shows the frequency of the three statements that were picked most frequently by the students in each category.

TABLE VIII

STATEMENTS THAT STUDENTS PICKED BECAUSE THEY DEEMED THEM TO BE SIMILAR TO THEIR OWN THINKING. SHOWN ARE THE THREE STATEMENTS MENTIONED MOST FREQUENTLY IN EACH CATEGORY. ITEMS ARE IDENTIFIED BY THEIR NUMBER AND THE FACTOR THEY ARE LOADING ON.

	1st year students		**3rd year students**	
… is most how I think.	N=303	No. 30/1 (17,5%) No. 9/1 (10,9%) No. 17/1 (8,9%)	N=144	No. 30/1 (31,9%) No. 17/1 (11,1%) No. 25/1 (10,4%)
… is second most how I think.	N=297	No. 22/1 (12,1%) No. 9/1 (9,8%) No. 24/1 (8,4%)	N=138	No. 9/1 (13,8%) No. 17/1 (12,3%) No. 25/1 (11,6%)
… is third most how I think.	N=289	No. 22/1 (14,2%) No. 30/1 (10%) No. 17/1 & No. 24/1 (9%)	N=135	No. 9/1 (11,9%) No. 22/1 (11,1%) No. 17/1 (11,6%)

There is a correlation between the students' RMI score and the factor that their most preferred item is loading on (Spearman's $r=-.018$, $p<.01$): As students' RMI score increases – indicating a more sophisticated level of reflective judgment – the factor number decreases (Factor 1 indicates the highest level of RJ while Factor 3 indicates the lowest level of RJ).

4. DISCUSSION

The objective of this study was to develop a test for assessing reflective judgment in medicine according to the Reflective Judgment Model (RJM). Hence, a pilot version of the Reasoning about Medical Issues Test (RMI) was administered to two cohorts of medical students in their first and third year. Our primary goal was to test the conceptual model of the RJM by means of a confirmatory factor analysis. Unfortunately, we were not able to achieve this goal. Neither of the three models tested with the CFA did show an acceptable fit. Several reasons come to mind that might explain this result. First of all, there is some agreement that epistemic cognitions are difficult to operationalize. Since items are rather short it is often challenging to capture the complex meaning of a construct such as reflective thinking as it is expressed in the RJM (Priemer 2006). Furthermore, this was the first attempt to construct a standardized test based on the RJM so the items have not been analyzed psychometrically prior to their use in the test. Thus, at least some of the items might not have been precise enough to meet the rather restrictive requirements of the CFA testing (Hurley et al. 1997). Another source of error or noise might be the mapping of the items to the levels and stages of the RJM that was used as the template for the CFA. The mapping was based on expert ratings. While all the experts had considerable experience in medical education as well as pedagogical psychology their knowledge and expertise regarding the RJM might have been limited producing too much vagueness in the classification of the items.

With regard to the results of the exploratory factor analysis the difficulties were less pronounced. It yielded a three factor solution allowing for a meaningful interpretation. Factor 1 (“complex reasoning”) captures the typical characteristics of medical knowledge and reasoning. Individuals agreeing with these items do acknowledge the inherent uncertainty in medical evidence as it is constantly evolving and developing. Yet, they are still confident that it is possible and necessary to evaluate the quality of arguments regarding their plausibility and their use of the available evidence. They recognize that individual differences do exist and explain them e.g., with different frames of reference. Factor 2 (“skeptic reasoning”) in contrast displays a different attitude. One might argue that individuals agreeing with these items seem to acknowledge that different opinions regarding medical issues do exist. But

instead of accepting this fact as an inevitable consequence of the uncertainty inherent in medical knowledge and problems they rather think that it roots in individual or collective bias, arbitrariness or even dishonesty. It seems as if the insight, that uncertainty and controversy regarding medical issues and problems do exist causes some discomfort, probably because these students do not know yet, how to handle these challenges. They do however trust official guidelines as the expression of collective wisdom and rigorous practice. Factor 3 ("experience-based reasoning") finally, conveys a rather simple understanding of epistemological issues as individual experience is the most prominent source of evidence and security here. While uncertainty might exist, it can be diminished or eliminated by referring to one's own experience or the experience of experts.

These descriptions suggest that the empirically derived factors do resemble the levels of the Reflective Judgment Model to a large extent. In fact, compared to the item mapping of the experts during the test construction process there is also considerable overlap. The empirical result is certainly no perfect match as two of the three factors do also embrace items that the experts allocated to other levels or stages. Four of the eleven items included in the complex reasoning factor were supposed to relate to stage five (quasireflective) and one to the prereflective level. However, the latter (No. 19: "Uncertainty with regard to medical decision is primarily a matter of insufficient experience and knowledge") has a negative loading which means that individuals exhibiting complex reasoning rather disagree with this statement which is in line with the thematic focus of this factor. Similarly, one of the seven items of the skeptical reasoning factor was supposed to relate to the reflective stage by the experts and has a negative loading (No. 6: "I trust the official medical guidelines because they are a result of a rigorous and systematic review process from experienced experts."). Again, this makes sense with regard to the skepticism reflected in this factor. This might also be true for item No. 13 ("If all experts were honest such controversies wouldn't exist") that was assigned to the prereflective level by the experts. Maybe the experts and the students interpreted this statement differently: The experts might have focused more on the second part ("such controversies wouldn't exist") which would rather exhibit a simplified understanding of the complexity of scientific discourse. The "skeptic" students on the other hand, might

have focused more on the first part (“if all experts were honest”) because cases of commercially biased experts or scientific misconduct have been repeatedly reported in the news. The experience-based reasoning factor however, does only include items that were supposed to relate to the lowest, the prereflective level. Thus, since the majority of the items constituting the three factors still come from their matching stage or from neighboring stages, the notion that the three levels represent a hierarchy of epistemological beliefs remains intact despite the differences in the empirical and the conceptual mapping of the items. Nevertheless, the empirical model is not as sophisticated as the original RJM and further studies will be necessary to find out whether revising some of the items or adding new items will eventually lead to a more differentiated model that reproduces not only the three levels, but also the seven stages as well as the nested structure of the RJM.

After adjusting the algorithm for calculating the preferred judgment level of the students to the three factor solution the resulting values were 2.19 for the first and 2.22 for the third year students, indicating that both groups mainly argued on the skeptic reasoning level (i.e. quasireflective). Arguing on that level means that the students have some awareness of different opinions and uncertainty as a frequent or typical component of medical problem solving. However, they still lack a deeper understanding on how to explain conflicting views other than by relating them to individual or idiosyncratic preferences (King and Kitchener, 1994). As mentioned above, studies measuring the reflective judgment level of medical students are rare, making it difficult to interpret these data. In the qualitative study with GP trainees in Belgium the detected level of reflective judgment was pre- or quasireflective which would be rather low if one assumes that there should be some development towards higher levels during medical education (Roex et al., 2009). In contrast, US dental students progressed from the quasireflective to the reflective level within their third year which seems to be a rather significant step in such a short period of time (Boyd, 2002). With an effect size of 0.19 the difference between the first and the third year students in our study is rather small and even the more advanced students argue primarily on the skeptic reasoning, i.e., quasireflective level. However, despite these rather subtle differences there is yet another indicator suggesting that some development between the first and third year might take place. Students in both

groups picked statement No. 30 (“While there is no such thing as a definite truth in medicine you can still assess the quality of a specific proposition or claim e.g. in terms of plausibility, transparency and credibility.”) most frequently as the one that is most similar to how they think. This statement loaded on the complex reasoning factor by the EFA and according to the prior expert rating it was assigned to the highest stage (7) of the RJM. However, as Table VIII reveals only 17.5% of the first-year students made this their first choice compared to 31.9% of the third-year students. Similar changes – less pronounced though – are noticeable for statement No. 17 and No. 25. Both items loaded on the complex reasoning factor and were assigned to stage 7 (17) and stage 6 (25) during the expert rating. Thus, it seems that the more advanced students pick less statements more frequently crystallizing the complex reasoning level more clearly. Nevertheless, further evidence especially from longitudinal studies will be needed in order to find out whether this interpretation holds true and students’ reflective judgment level really increases over the course of their studies.

To further interpret such findings, it is of utmost importance to refer to the content of the respective curriculum. The six-year curriculum at the institution of our study for instance, is rather “traditional”. During the first two pre-clinical years didactic lectures and other rather teacher-centred formats prevail conveying mainly basic science content. Students have only very little opportunity to develop critical or reflective thinking as their major concern is to master the huge amount of facts. As representative national surveys have shown, this is a general problem affecting medical education all over Germany (Vöttiner and Woisch, 2010). Thus, the finding that the difference between first and third year students in our study is rather subtle might reflect the fact that students’ reflective judgment has not (yet) been challenged or specifically addressed during the course of their studies.

4.1. Limitations

Overall, the empirical model derived in our study is limited. The internal consistency of the scales is quite low as is the discrimination power of some of their items. Thus, in its current form the RMI-test is certainly not suited for routine application in medical education let alone for statements about an individual person. As already mentioned, our link between the conceptual model and the template for

the CFA might have been too vague. Despite the fact that the levels and stages of the RJM are defined by qualitative attributes, the stages of one level also share some commonalities. Thus, it might be rather difficult to capture especially the characteristics of the stages with sufficient precision. Furthermore, our expert panel might have been too small to reach a more reliable consensus on the assignment of the items. The experts might also need a more comprehensive instruction and preparation to assign the items with greater certainty to the different stages of the RJM. The participants of our sample were medical students from just one institution which might have resulted in groups that are too homogenous in terms of their responses. Thus, it would have been desirable to have a sample exhibiting the spectrum of reflective judgment more fully.

4.2. Conclusions

Overall, our findings paint a rather mixed picture. As others have noticed before, it seems rather difficult to develop a meaningful quantitative measure for epistemological beliefs. We did not succeed either in our attempt to reproduce the sophisticated structure of the RJM directly. However, our findings suggest that basic assumptions of the RJM seem to be valid. Firstly, the distribution of the RMI scores in both groups indicates that individuals argue on different levels of complexity and differentiation when faced with ill-defined problems. Secondly, the differences between the first- and the third-year students alludes to the developmental aspect of the RJM as students' reflective judgment seems to evolve over the course of their studies. Further evidence will be needed to corroborate the quality of our instrument and to substantiate these assumptions in order to better understand how medical education can foster the development of students' reflective judgment.

CITED LITERATURE

Abraham, I., Killackey-Jones, B. (2002). Lack of evidence-based research for idiopathic low back pain: the importance of a specific diagnosis. Archives of Internal Medicine, 162, 1442-1444

Blaha, M. J., Nasir, K., Blumenthal, R. S. (2012). Statin therapy for healthy men identified as "increased risk". JAMA, 307, 1489-1490

Boyd, L. D. (2002). Reflections on clinical practice by first-year dental students: a qualitative study. Journal of Dental Education, 66, 710-720

Broeders, M., Moss, S., Nyström, L., Njor, S., Jonsson, H., Paap, E., Massat, N., Duffy, S., Lnge, E., Paci, E. (2012). The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. Journal of Medical Screening, 19(1_suppl), 14-25

Buehl, M. M., Alexander, P. A. (2006). Examining the dual nature of epistemological beliefs. International Journal of Educational Research, 45, 28-42

Cassel, C. K., Guest, J. A. (2012). Choosing wisely: helping physicians and patients make smart decisions about their care. JAMA, 307, 1801-1802

Cholesterol Treatment Trialists (CTT) Collaborators (2012). The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. The Lancet 380: 581-590

DeBacker, T. K., Crowson, H. M., Beesley, A. D., Thoma, S. J., Hestevold, N. L. (2008). The challenge of measuring epistemic beliefs: An analysis of three self-report instruments. The Journal of Experimental Education, 76(3), 281-312

Dewey, J. (1910). How we think. (Boston: D.C.Heath and Co.)

Deyo, R. A. (2002). Diagnostic evaluation of LBP: reaching a specific diagnosis is often impossible. Archives of Internal Medicine, 162, 1444-1447

Eastwood, J. L., Koppelman-White, E., Mi, M., Wasserman, J. A., Krug III, E. F., Joyce, B. (2017). Epistemic cognition in medical education: a literature review. International Journal of Medical Education, 8, 1

Elstein, A. S., Shulman, L. S., Sprafka, S. A. (1978). Medical problem solving. An analysis of clinical reasoning. (Cambridge, MA: Harvard University Press)

Eraut, M. (1985). Knowledge creation and knowledge use in professional contexts. Studies in Higher Education, 10, 117-133

Fürntratt E (1969). [Determining the number of interpretable common factors in factor analyses of psychological data] Zur Bestimmung der Anzahl interpretierbarer gemeinsamer Faktoren in Faktorenanalysen psychologischer Daten. Diagnostica, 15, 62-75.

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. Educational Psychology Review, 13(4), 353-383

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. Educational Psychologist, 39, 43-55

Hofer, B. K., Bendixen, L. D. (2012). Personal epistemology: Theory, research, and future directions. APA educational psychology handbook, 1, 227-256

Hofer, B. K., Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67, 88-140

Hunnink, M. M., Weinstein, M. C., Wittenberg, E., Drummond, M. F., Pliskin, J. S., Wong, J. B., Glasziou, P. P. (2014). Decision making in health and medicine: integrating evidence and values. (Cambridge: Cambridge University Press)

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. Journal of Organizational Behavior, 18, 667-683

Ilgen, J. S., Eva, K. W., de Bruin, A., Cook, D. A., Regehr, G. (2018). Comfort with uncertainty: reframing our conceptions of how clinicians navigate complex clinical situations. Advances in Health Sciences Education, 1-13

King, P. M., & Kitchener, K. S. (1994). Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults. (San Francisco, CA: Jossey-Bass)

King, P. M., Kitchener, K. S. (2004). Reflective judgment: Theory and Research on the development of epistemic assumptions through adulthood. Educational Psychologist, 39, 5-18

Kitchener, K. S., King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. Journal of Applied Developmental Psychology, 2, 89-116

Maher, C., Underwood M., Buchbinder, R. (2017). Non-specific low back pain. The Lancet, 389, 736-747

Malterud, K., Guassora, A. D., Reventlow, S., Jutel, A. (2017). Embracing uncertainty to advance diagnosis in general practice. The British Journal of General Practice, 67, 244-245

Mann, K., Gordon, J., MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. Advances in Health Sciences Education, 14, 595

Montgomery, K. (2006). How doctors think. Clinical judgment and the practice of medicine. (Oxford: Oxford University Press.)

Moser, P. K. (2002). Introduction. (In P. K. Moser (Ed.) The Oxford Handbook of Epistemology (pp. 3-24). Oxford: Oxford University Press.)

Muis, K. R., Bendixen, L. D., Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. Educational Psychology Review, 18, 3-54

Ng, S. L., Kinsella, E. A., Friesen, F., Hodges, B. (2015). Reclaiming a theoretical orientation to reflection in medical education research: a critical narrative review. Medical education, 49, 461-475

Norman, G. (2005). Research in clinical reasoning: past history and current trends. Medical education, 39, 418-427

Redberg, R. F., Katz, M. H. (2012). Healthy men should not take statins. JAMA, 307, 1491-1492

Roex, A., Clarebout, G., Dory, V., Degryse, J. (2009). Can ill-structured problems reveal beliefs about medical knowledge and knowing? A focus-group approach. BMC Medical Education, 9, 62

Sandoval, W. A., Greene, J. A., Bråten, I. (2016). Understanding and promoting thinking about knowledge: Origins, issues, and future directions of research on epistemic cognition. Review of Research in Education, 40, 457-496

Sullivan, W. M., Rosin, M. S. (2008). A new agenda for higher education. Shaping a Life of the mind for practice. (San Francisco, CA: Jossey-Bass.)

Vöttiner, A., Woisch, A. (2012). [Study Quality Monitor 2010. Study Quality and Study Conditions] Studienqualitätsmonitor 2010. Studienqualität und Studienbedingungen. HIS:Forum Hochschule, 4, 2012

Wood, P. K. (1997). A secondary analysis of claims regarding the reflective judgment interview: internal consistency, sequentiality and intra-individual differences in iII-structured problem solving. Higher Education, 12, 243-312

Wood, P., Kitchener, K., Jensen, L. (2002). Considerations in the design and evaluation of a paper-and-pencil measure of epistemic cognition. Personal epistemology: The psychology of beliefs about knowledge and knowing, 1, 277-294.

(APPENDICES)

Dr. G. Fabry
Albert-Ludwigs-University
Medical Psychology & Sociology
Rheinstr. 12 / 79104 Freiburg / Germany

Phone : ++49 / +761 / 203 – 5512
Fax: ++49 / +761 / 203 – 5514
fabry@uni-freiburg.de

Dear students,

today we would like to invite you to participate in a research study conducted by Dr. G. Fabry of Albert-Ludwigs-University Freiburg, Department of Medical Psychology & Sociology. We would like learn what you and other medical students think about the handling of controversial issues in medicine.

Your participation in this study is voluntary, nothing you say on the questionnaire will in any way influence your present or future course of study. You can stop at any time. You do not have to answer any questions you do not want to answer. Refusal to take part in or withdrawing from this study will involve no penalty or loss of benefits you would receive otherwise. Your participation in this study does not involve any physical or emotional risk to you beyond that of everyday life. Your responses will strictly be handled anonymously, no one will be able to identify you.

If you consent to participate in the study please proceed.

We plan to survey you again later in your studies. Thereby, we want to find out whether your ratings might have changed. To match your current responses with those from future surveys we need a code that enables us to relate your responses without identifying you personally.

Thus, you should be the only one who knows this **code** and you should definitely remember it.

To ensure this please note in the respective boxes:

1.	the first letter of **your mother's first name**	☐	(e.g. **E** für **E**ve)
2.	the first letter of **your mother's maiden name**.	☐	(e.g. **M** für Miller)
3.	your mother's birthday (**just the day, two digits**)	☐☐	(e.g. **05** für April **5th,** 1958)
4.	the **month of your mother's birthday** (**two digits**)	☐☐	(e.g. **04** für **April** 5th, 1958)

Please provide us with the following information on your person:

Are you… female O male O?

How old are you? …….. years

For how many semesters have you been studying medicine? ..... (1-12)

Before you started medical school: Did you complete a training in a health profession (e.g. nursing, paramedic) and/or did you work in health care elsewhere (e.g. temporary job, assistant)?

yes O no O

If your answer is yes: For how long have you been working? …….. years

On the following pages you will firstly be provided with an introductory explanation and then with a couple of statements that we would like you to review. The focus is not on specialized medical knowledge but the way arguments are used.

Controversial issues in Medicine

Quite often, questions regarding medical issues are discussed intensively and controversially e.g. when and how to treat a patient with a certain disease or a person with a risk factor or whether or not a screening test for cancer should be recommended.

A good example is the use of so called statins for the treatment of elevated cholesterol levels. Although there is no doubt that elevated cholesterol levels are associated with a greater risk of cardiovascular diseases, there is ongoing controversy about when and how to treat otherwise healthy individuals with elevated cholesterol levels. Some experts argue that in addition to dietary modification and increased physical activity these individuals should be treated with a statin (a drug that lowers the cholesterol level). Thus, the risk for severe consequences such as myocardial infarction and stroke could be decreased. In contrast, other experts argue that such a drug therapy is not indicated because its potential side effects (e.g. muscle weakness, fatigue, memory loss and even diabetes) are so severe that they outweigh the potential benefits (JAMA 307/14:1489-1492, 2012).

Another example is screening for breast cancer. Breast cancer is the most common cancer in women in industrialized countries. Five years after the diagnosis about 89% of all breast cancer patients are still alive; however, among women between the age of 30 and 60 breast cancer is the leading cause of death. The controversy concerns the risks and benefits of breast cancer screening. Proponents of breast cancer screening argue that an early diagnosis will open up more therapeutic options resulting in more definite remissions and thus in a reduction of breast cancer mortality. On the other hand critics claim that although breast cancer screening does indeed lead to more and earlier diagnoses, the mortality does not decline as much as expected. Since breast cancer mortality does also decline in countries or regions where no comprehensive screening is implemented, some experts think that other factors such as improved therapies are more important to reduce breast cancer mortality. Furthermore, with comprehensive breast cancer screening several women will unnecessarily be diagnosed with breast cancer (and undergo treatment) because their cancer would otherwise not have become apparent. Even more women will undergo an invasive biopsy that will turn out to be normal. (Source: J Med Screen 19 Suppl 1: 14-25; Dtsch Arzteblatt 105: 131-6)

Obviously, even expert might have different positions or opinions regarding such problems. This brings up at least two questions:

1) Why do different opinions exist at all on medical issues?
2) Are the justifications given for the different opinions equally convincing?

On the following pages you will find a number of statements that a person might use

- to justify why he or she takes a certain position with regard to a medical issue such as using statins to lower high cholesterol levels or recommending breast cancer screening;
- to explain why different individuals hold different opinions on these issues.

Please indicate for each statement how appropriate you think it is **to justify one's own point of view or to explain the diversity of opinions.** For each statement you can use a 5-point scale that ranges from "very inappropriate" to "fully appropriate". When reviewing the statements you might think of a person making such a statements in a discussion to confirm his or her opinion or to explain why different opinions exist at all.

Please indicate for each statement how appropriate it is to confirm an opinion or to explain the diversity of different opinions.	**very inappropriate**				**fully appropriate**
1. I have treated many patients with statins and never witnessed any severe side-effects.	○	○	○	○	○
2. Certainly, other individuals might come to alternative conclusions if they pick the "right" arguments for them.	○	○	○	○	○
3. Such controversies arise because we do not know enough yet. Future evidence will prove who is right.	○	○	○	○	○
4. I don't trust results from drug studies because they are sponsored by the pharmaceutical industry and thus, the outcomes are biased in favor of the respective manufacturer.	○	○	○	○	○
5. The major reason for the different opinions is that some experts present the evidence from studies in a biased way in favor of specific interests.	○	○	○	○	○
6. I trust the official medical guidelines because they are a result of a rigorous and systematic review process from experienced experts.	○	○	○	○	○
7. Faced with multifaceted findings from scientific studies you have to arrive at your own conclusions and since people are different their interpretations are different too.	○	○	○	○	○
8. Before I become too confused by controversial findings from studies I prefer to rely on my personal experience with patients.	○	○	○	○	○
9. Even experts hold different opinions as they evaluate the available evidence differently. However, some conclusions are more appropriate than others and reflect a more comprehensive synthesis of the available information.	○	○	○	○	○
10. Different opinions exist because we do not know enough yet. Thus, the whole thing is just a matter of belief.	○	○	○	○	○
11. On a recent conference Professor X who is a renowned expert in the field advocated statin therapy very much. I will follow this advice.	○	○	○	○	○
12. It all depends on the perspective you take because this determines how e.g. the results from different studies are interpreted and used to build an argument.	○	○	○	○	○
13. If all experts were honest such controversies wouldn't exist.	○	○	○	○	○
14. Different positions arise since knowledge always has to be constructed on the basis of theories and data.	○	○	○	○	○
15. Presumably, we will never agree on such issues as the scientific evidence is constantly evolving and changing.	○	○	○	○	○
16. Medical science is not just based on facts. Personal views and opinions do also matter resulting in different points of view.	○	○	○	○	○
17. Since definite answers in medicine are rare one must evaluate a certain viewpoint e.g. with regard to how plausible the argumentation is based on current evidence.	○	○	○	○	○
18. I have personally seen patients dying whose life could have been saved by a timely therapy. Thus, I am in favor of screening tests.	○	○	○	○	○
19. Uncertainty with regard to medical decisions is primarily a matter of insufficient experience and knowledge.	○	○	○	○	○

Please indicate for each statement how appropriate it is to confirm an opinion or to explain the diversity of different opinions.	**very inappropriate**		.....		**fully appropriate**
20. It is very difficult to know something for sure. Eventually you can be certain enough to form a personal opinion on a specific issue.	○	○	○	○	○
21. Since there is so much evidence on medical issues it is quite arbitrary what one uses to justify a certain position.	○	○	○	○	○
22. No matter how competent, experienced or knowledgeable you are: uncertainty regarding medical evidence will never completely dissolve.	○	○	○	○	○
23. I rely on medical textbooks very much because they are written by experts in the field.	○	○	○	○	○
24. The different viewpoints are caused by different perspectives: A biochemist takes a different perspective than a sociologist.	○	○	○	○	○
25. Knowledge in general is fluctuating and prone to uncertainty.	○	○	○	○	○
26. I once had to take statins myself and had a horrible experience. Since then I am very restrictive in prescribing these drugs.	○	○	○	○	○
27. When there seems to be more than one answer I usually adhere to what I learned during my residency in XY which so far has never failed me.	○	○	○	○	○
28. As long as we don't know the definite answer everybody can bend the evidence to make it fit for his or her own purpose.	○	○	○	○	○
29. It's in the nature of medical problems that they don't have a definite answer.	○	○	○	○	○
30. While there is no such thing as a definite truth in medicine you can still assess the quality of a specific proposition or claim e.g. in terms of plausibility, transparency and credibility.	○	○	○	○	○
31. Unfortunately, many experts are not credible because they don't act independently from third party or commercial interests.	○	○	○	○	○

In addition: Please pick from the above statements **up to three** that are most similar to your thinking. Please write down the respective number in the boxes below:

Statement…		
No.	**No.**	**No.**
… is **most** how I think	… is **second most** how I think	… is **third most** how I think.

Thank you very much indeed for your support and effort!

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Granted

September 7, 2016

Goetz Fabry, MD
UIC Medical Education Student
Freiburg University Medical School
Department of Medical Psychology
Rheinstrasse 12, 79104
Freiburg Germany

RE: Research Protocol # 2016-0839
"Reflective Judgement in Medicine"

Sponsors: None

Dear Dr. Fabry:

Your Claim of Exemption was reviewed on September 7, 2016 and it was determined that your research protocol meets the criteria for exemption as defined in the U. S. Department of Health and Human Services Regulations for the Protection of Human Subjects [(45 CFR 46.101(b)]. You may now begin your research.

UIC Exemption Period:	**September 7, 2016 – September 7, 2019**
Lead Performance Site:	**Freiburg University Medical School (Freiburg Germany)**
Other Performance Site(s):	UIC
Subject Population:	Adult (18+ years) subjects only
Number of Subjects:	FUMS: 700; UIC: 0; Total = 700

The specific exemption categories under 45 CFR 46.101(b) are:
(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices. such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods; and

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

You are reminded that investigators whose research involving human subjects is determined to be exempt from the federal regulations for the protection of human subjects still have responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.

2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.

3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).

4. Information for Human Subjects UIC Policy requires investigators to provide information about the research to subjects and to obtain their permission prior to their participating in the research. The information about the research should be presented to subjects as detailed in the research protocol and application utilizing the approved recruitment and consent process and documents.

Please be sure to use your research protocol number (listed above) on any documents or correspondence with the IRB concerning your research protocol.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact me at (312) 355-2908 or the OPRS office at (312) 996-1711.

Sincerely,
Charles W. Hoehne, B.S., C.I.P.
Assistant Director, IRB #7
Office for the Protection of Research Subjects

cc: Ilene Harris, Medical Education, M/C 591
Alan Schwartz, Medical Education, M/C 591

Ethik-Kommission ▫ Engelberger Straße 21▫ 79106 Freiburg

Herrn
Dr. med. Götz Fabry
Albert-Ludwigs-Universität Freiburg
Medizinische Fakultät
Medizinische Psychologie & Soziologie
Rheinstr. 12
79104 Freiburg

Internal mail

Albert-Ludwigs-Universität Freiburg

Ethik-Kommission

Prof. Dr. S. Pollak
(Vorsitzender)
M. Schmidt
(Komm. Geschäftsführer)

Engelberger Straße 21
79106 Freiburg

Tel. 0761/270-72600 / -72500
Fax 0761/270-72630

www.ethik-kommission.uniklinik-freiburg.de

Bearbeitet von:
Kristina Wehrle
Durchwahl: -72600
kristina.wehrle@uniklinik-freiburg.de

Freiburg, 03. Dezember 2015

Application No.:	**502/15** *(bitte stets angeben)*
Projectleader:	Dr. med. Götz Fabry Albert-Ludwigs-Universität Freiburg Medizinische Fakultät Medizinische Psychologie & Soziologie
Research Project:	Reflektiertes Urteilen in der Medizin [Reflective Judgement in Medicine]

Englisch translation of the vote of December 03, 2015

VOTE

Dear Dr. Fabry,

The Ethics Committee has reviewed your application specified above on the basis of the documents indicated in the Appendix according to Section 15 of the Code of Conduct of the General Medical Council of Baden-Württemberg. It does not have any ethical or legal objections to the conduct of the research project.

All changes in the study protocol and updates of the information sheet for patients and study participants resulting from such changes have to be submitted to the Ethics Committee for approval before putting them into effect.

The Ethics Committee would like to emphasize that the ethical and legal responsibility for the conduct of the research project lies with the project leader and the physicians participating in the study.

The Ethics Committee would like to point out that this vote is only valid for the persons indicted in the protocol / list in the application to the Ethics Committee who are involved in the research project and are members of Albert Ludwig University Freiburg. Thus physicians have also fulfilled their professional obligation to seek advice as stipulated in Section 30 subs. 4 Heilberufe-Kammergesetz (Professional Chamber Act of Healthcare Professions) and Section 15 subs. 1 of the Code of Conduct of the Baden-Württemberg General Medical Council.

Please inform the Ethics Committee in writing when the research project is completed or in case that it is prematurely terminated and submit a final report containing a summary of the results and conclusions of the study.

Forschungsvorhaben: 502/15

The Ethics Committee strongly recommends registering the research project in an open access register meeting the requirements defined by the World Health Organization (WHO). We may further point to the resolution of the Board of the Medical Faculty dated 24 April 2012 which stated i. a. that all clinical studies conducted within the scope of the Medical Faculty and the Freiburg University Medical Centre have to be registered in a WHO registry by the respective coordinating investigator. The DRKS (Deutsches Register Klinischer Studien – German Register of Clinical Trials) is recognized as a primary WHO registry and thus meets the requirements of the International Committee of Medical Journal Editors (ICMJE). For further information and registration go to *www.germanctr.de.* Unless already done, please let us know the respective registration number.

Yours sincerely,

Prof. Dr. S. Pollak
Chairman of the Ethics Committee

Appendix to our vote of December 03, 2015

List of submitted documents:

Received Nov 28, 2015

- Cover letter of Oct 26, 2015
- Application form, signed Oct 20, 2015
- Protocol, Version 7.1, Oct 16, 2015
- Information for Participants and Informed Consent
- Questionnaire, mpr_2015_1 (rmi7.1)
- CV Dr. Götz Fabry, Oct 19, 2015

Received Dec 02, 2015

- Cover letter of Nov 30,2015
- Application form (revieced)
- Protocol, Version 7.1, Oct 16, 2015
- Questionnaire „students", mpr_2015_1 (rmi7.1)
- Questionnaire „doctors", mpr_2015_1 (rmi7.1a)
- Recruiting advertisment

VITA

NAME: Götz Fabry

EDUCATION: M.D., Albert-Ludwigs-University, Freiburg, Germany, 1996

M.S., Health Professions Education, University of Illinois at Chicago, 2019

CURRENT POSITION: Assistant Professor, Department of Medical Psychology and Medical Sociology, Albert-Ludwigs-University, Faculty of Medicine, Freiburg, Germany

PROFESSIONAL MEMBERSHIPS: German Association for Medical Edcuation (GMA)

German Association for Medical Psychology (DGMP)

Association for Medical Education in Europe (AMEE)

Association for the Study of Medical Education (ASME)

AWARDS: Teaching Award, Albert-Ludwigs-University, Freiburg, 2018

Medical Teacher Poster Prize, Association for Medical Education in Europe (AMEE), 2015

Excellence in Reviewing Award, Academic Medicine, 2011

Instructional Development Award, Albert-Ludwigs-University, Freiburg, 2010

SELECTED PUBLICATIONS:

Gaupp R, Fabry G, Körner M: Self-regulated learning and critical reflection in an e-learning on patient safety for third-year medical students. International Journal of Medical Education, 2018; 9: 189-194

Schubach F, Goos M, Fabry G, Vach W, Boeker M: Virtual patients in the acquisition of clinical reasoning skills: does presentation mode matter? A quasi-randomized controlled trial. Bmc Med Educ, 2017; 17 (1) : 165

Gaupp R, Körner M, Fabry G: Effects of a case-based interactive e-learning course on knowledge and attitudes about patient safety: a quasi-experimental study with third-year medical students. Bmc Med Educ, 2016; 16: 172

Woelber JP, Spann-Aloge N, Hanna G, Fabry G, Frick K, Brueck R, Jahne A, Vach K, Ratka-Kruger P: Training of Dental Professionals in Motivational Interviewing can Heighten Interdental Cleaning Self-Efficacy in Periodontal Patients. Front Psychol, 2016; 7: 254