

Challenging the Construct Validity of Mindfulness Assessment—a Cognitive Interview Study of the Freiburg Mindfulness Inventory

Florian Belzer · Stefan Schmidt · Gabriele Lucius-Hoene ·
Johann F. Schneider · Claudia L. Orellana-Rios ·
Sebastian Sauer

Published online: 27 October 2012
© Springer Science+Business Media New York 2012

Abstract The Freiburg Mindfulness Inventory (FMI) is a widely used questionnaire of self-reported mindfulness. However, doubts have been expressed as to whether an adequate comprehension of the items of the FMI is independent of one's mindfulness experience (ME). The aim of the present study was to determine with qualitative methods whether and how ME influences the response to the FMI items. Two groups, matched for gender, education, and age ($N=11$ each), with and without mindfulness training, completed the FMI while at the same time applying the technique of thinking aloud. The protocols of the two samples were compared using three different strategies: (1) predefined criteria on the comprehension of each item developed by FMI experts, (2) a coding scheme developed to identify differences in specific cognitive processes, and (3) qualitative analysis of comprehension patterns. The results showed that (1) participants with ME fulfilled the item criteria for comprehension much more than participants without ME. (2) The coding scheme demonstrated greater *comprehension*

difficulties in the sample without ME. Differences in *judgment* processes between groups could not be found. (3) Qualitative analysis revealed comprehension problem patterns especially for eight items for the comparison group. It is concluded that a modification of the wording of several FMI items is necessary and that there is insufficient construct validity to use the current FMI in mindfulness-naïve samples. This may also be true for other scales tapping into the assessment of the awareness component of mindfulness, and it is recommended to also check their construct validity.

Keywords Freiburg Mindfulness Inventory (FMI) · Construct validity · Think-aloud technique · Mindfulness assessment · Qualitative analysis · Mixed methods

Introduction

Despite the rapidly growing literature concerning the effectiveness of mindfulness interventions in the past 30 years, the rise in the measurement of self-rated mindfulness has been a relatively recent development. According to Bear (2007), there is an urgent need for reliable and valid tools to measure mindfulness, because otherwise it is not possible (1) to determine if the proposed skills of a mindfulness-based intervention have actually been acquired, (2) to explore and describe potential action mechanisms of mindfulness-based interventions, and (3) to differentiate specific mindfulness skills from other possible effects of mindfulness trainings such as social support.

Self-report measures are the most common methods for the measurement of mindfulness. The reason for the lack of more objective measures is partly due to the subjective nature of mindfulness. Currently, 11 questionnaires claiming to measure mindfulness have been published: the Development

F. Belzer · S. Schmidt (✉) · C. L. Orellana-Rios
Center for Meditation, Mindfulness and Neuroscience Research,
University Medical Center Freiburg,
Breisacher-Str. 115b,
79106 Freiburg, Germany
e-mail: stefan.schmidt@uniklinik-freiburg.de

G. Lucius-Hoene
Institute of Psychology, University of Freiburg,
Freiburg, Germany

J. F. Schneider
Institute of Psychology, University of Saarland,
Saarbrücken, Germany

S. Sauer
Generation Research Program, Human Science Center,
Ludwig-Maximilians-University,
Munich, Germany

Mindfulness Survey (Solloway and Fisher 2007), the Effects of Meditation Scale (Reavley and Pallant 2009), the Cognitive and Affective Mindfulness Scale Revised (Feldman et al. 2007), the Toronto Mindfulness Scale (Lau et al. 2006), the Kentucky Mindfulness Scale (Baer et al. 2004), the Mindfulness Attention and Awareness Scale (Brown and Ryan 2003), the Philadelphia Mindfulness Scale (Cardaciotto et al. 2008), the Five Facets Mindfulness Questionnaire (Baer et al. 2006), the Southampton Mindfulness Scale (Chadwick et al. 2008), the Freiburg Mindfulness Inventory (FMI, Walach et al. 2006), and the Langer Mindfulness/Mindlessness Scale (Haigh et al. 2011). The Freiburg Mindfulness Inventory is an instrument with explicit reference to the Buddhist roots of the mindfulness concept—insofar as experienced long-term Buddhist teachers rated items for the final questionnaire for suitability to capture the concept of mindfulness. The FMI Short Version used in this study consists of 14 items and is rated on a 4-point Likert scale with answer options ranging from 1 (*rarely*) to 4 (*almost always*). The FMI Short Version is intended to measure a general factor of mindfulness and is highly correlated ($r=0.95$) with the long form.

Measuring mindfulness with a self-report questionnaire poses a number of difficulties that require deeper exploration. First of all, one can argue that the experience of mindfulness is paraconceptual (Marcel 2003) or preverbal (Fattah 2009) in nature. There have been doubts expressed about whether it is possible to formulate the experience of mindfulness in a self-report item (Grossman 2008). It has been postulated that there is both an epistemological and a methodological problem in verbalizing mindfulness experience (Fattah 2009). Grossman (2008) mentions a number of difficulties concerning mindfulness self-report questionnaires; these are among others the following: (1) conceptual difficulties and disagreement even among experts about the meaning of “mindfulness,” (2) discrepancies between the self-estimated and the “true” amount of mindfulness experience among respondents, and (3) a different semantic interpretation of questionnaire items among respondents, partly due to a different extent of meditation experience. These statements are in general debatable and have so far not been subject to empirical investigations. However, if true, especially the third issue, they would represent serious problems in measuring the concept of mindfulness. It is easily conceivable that participants of a Mindfulness-Based Stress Reduction (MBSR) intervention who never meditated before have a totally different concept of mindfulness than highly experienced meditators. It is also clear that their mindfulness concept could have been modified by the MBSR course. Therefore, it cannot be readily determined if the response to the items of an inventory like the FMI really detects an improvement of the mindful practice or merely reflects a change in the cognitive concept respective to the internal reference framework of mindfulness.

This problem is well studied in quality of life research and has been termed *response shift* (Güthlin 2004).

Another problem is the poorly described process of item construction with respect to *item comprehension* in the literature. Often researchers have a strong emphasis on the evaluation of *item suitability*, which in the case of mindfulness questionnaires should be conducted by experts of the matter (e.g., Buddhism teachers evaluated the suitability in the case of FMI, Walach et al. 2006), as well as on the *psychometric properties* of the items. But for an acceptable construct validity, one also needs a well-grounded expertise in order to select adequate items for a questionnaire with respect to comprehension problems (e.g., misunderstanding, ambiguity, different comprehension than intended). Many publications of psychometric scales lack information regarding this aspect. Given the high practical relevance and use of mindfulness-based interventions, the need for valid mindfulness scales arises. This need justifies further research efforts to address the present difficulties in conceptualizing and measuring mindfulness. More thorough investigation into the semantic understanding of respondents towards mindfulness items may be one avenue to improve the quality of measurement instruments in the field. Qualitative methods may provide an effective way to deepen our understanding of the comprehension patterns of respondents to mindfulness questionnaires. To date, such approaches have not been conducted.

In this study, we followed an explorative and qualitative approach. The cognitive interviewing technique of thinking aloud was employed to obtain data. Cognitive interviews within survey research describe methods used to thoroughly examine a particular questionnaire's item. The methods are intended to be additional to and not a substitution for psychometric item analysis. In short, they attempt to take a look into the cognitive processes of an individual between stimulus (question) and response (check mark). According to Tourangeau (1984), the main cognitive processes underlying questionnaire completion are *comprehension, retrieval, judgment, formatting, and verification*. Without building a strict temporal sequence, these processes are said to be part of every response to a questionnaire item (Sudman and Bradburn 1982).

To access the content of cognitive processes, we used the “thinking aloud” technique. This technique asks participants to verbalize their thoughts during a given task. Because thoughts are not audible, a strict definition would be “to speak aloud what one is thinking right now” (Hak et al. 2008). In survey research, the goal of thinking aloud is to gain deepened knowledge about the respondent's comprehension of the item and of the judgment processes which underlie their response (Prüfer and Rexroth 2005). The thinking aloud method may reveal previously unknown difficulties in relation to questionnaire items. This is an advantage over other more quantitative survey research

methods which are deductive in nature and thus require hypotheses of possible item problems (Prüfer and Rexroth 2005). For more detailed information about the technique of thinking aloud and cognitive interviews, see the work of Huber and Mandl (1982), Ericsson and Simon (1993), Presser et al. (2004), Prüfer and Rexroth (2005), and Willis (2005).

The aim of the present study was to answer the following questions:

1. Does adequate comprehension of the FMI items depend on the level of proficiency or knowledge of mindfulness?
2. Do respondents with or without prior mindfulness experience differ in their cognitive processes as verbalized during the administration of the FMI?
3. What exactly is the difference in comprehension between mindfulness practitioners and non-meditators?

Method

Participants

Two groups were compared, mindfulness practitioners (mindfulness experience group, MEG) and non-meditators who had never practiced any form of mindfulness-based training (comparison group, CG). Meditators had at least a regularly weekly meditation practice for more than 4 years. Both groups were matched regarding age (± 5 years), gender, and educational background. A total of 22 participants, 11 per group, were recruited through newspaper ads and public notices. Both groups were almost equally divided between male (five) and female (six), and all participants but one in each group had A-levels/college entry level degrees; see Table 1 for details.

Procedure

Cognitive Interviews Participants were invited to a single-session sitting. To become familiar with the task of thinking aloud, every session started with an introduction in which each participant was asked to verbalize thoughts while observing a picture for 2 min. The picture contained easily

describable everyday scenes (Schneider and Reichl 2006). This procedure allows for the estimation of verbal fluency of the person, i.e., the capability to speak fluently as measured by the amount of words—possibly an indicator of the ability to think aloud (Schneider and Reichl 2006). The spoken words of both groups were counted and compared.

A validated manual for conducting cognitive interviews (“Three-Step Test-Interview”; Hak et al. 2008) was used to set the interview schedule. The schedule consisted of three successive steps. The first step was the actual procedure in which participants thought aloud while answering the 14 items of the FMI short version. The second step consisted of a focused interview in which the interviewer debriefed participants after step 1 and asked them to think aloud retrospectively in case the concurrent think-aloud task failed (e.g., when participants showed only very limited verbalizations). The third step consisted of a semi-structured interview in which the interviewer asked additional questions such as, for example, how the participant comprehended a particular item or term (verbal probes). All verbalizations were recorded with a digital voice recorder and transcribed without using specific transcription rules (Kuckartz 2006).

Analysis of Data Evaluation of the verbal protocols consisted of three major strategies: first, the elaboration of external criteria for each item of the FMI by its authors; second, the development of a coding scheme in accordance with Chi (1997) and Brown et al. (2009); and third, a basic qualitative analysis method following Mayring (1983, 2002). Each step is described in detail below.

Development of External Criteria for Each Item The development of external criteria for each item consisted of a description of the specific mindfulness facet reflected by the item. For this, two of the developers of the FMI (Harald Walach and Stefan Schmidt) defined the items and set indications for an adequate item comprehension. An example can be found in Table 2.

Verbal protocols of all participants were evaluated for matching these criteria. Criteria were considered as matched if at least one indication for an adequate item comprehension was fulfilled. If so, the verbal protocol was assigned category I (*mindful comprehension*). Category II (*no item criteria fulfilled*) was assigned where none of the indications were

Table 1 Descriptive statistics of both groups

	<i>n</i>	Sex f/m	Age M (SD)	Meditation experience M (SD)
MEG	11	5/6	48.7 (10.01)	15.9 (12,35)
CG	11	5/6	47.0 (9.95)	–

Age and meditation experience in years

Sex: *f* female, *m* male

Table 2 Example of external item criteria for FMI-Item 10

Item 10. I watch my feelings without getting lost in them.
An adequate comprehension of this item contains the understanding of one of the following criteria:
1. Awareness facet of mindfulness
2. Mindfulness of emotion
3. Disidentification with emotion

Table 3 Categories of the developed coding system and examples of categorized verbal protocol segments

Cognitive process	Content	Example
Comprehension	Explicit C. problems	"I do not understand the question!"
	Implicit C. problems	"How am I supposed to pay attention to what's behind my actions?"
	Criticizes item	"Oh, this sounds weird!"
	Defining item	"This item means that one should be good to himself."
	Consideration	"Let me think about it, hmm..."
	Restates item	Participant rereads the item.
	Pause	Participant is silent for at least 6 s.
Judgment	Personal experience	"That is exactly what I try to do in my meditation."
	Self, descriptive	"I would say I am a person who is indeed very open."
	Self, prescriptive	"That is something I should do more."
	General statement	"I think in general people do not have the time to do this."
	Specifying, simple	"Hmm, it depends... sometimes yes, sometimes no."
	Self, specifying, complex	"If I am feeling fine and the weather is good, I can do this."
	Agreement	"Yes, I think so."
	Disagreement	"No, definitely not."
	Positive item assessment	"This is a nice question."
	Negative item assessment	"I do not like this item."
Response	"Almost always"	4 (Four-step Likert scale point 4)
	"Fairly often"	3 (Four-step Likert scale point 3)
	"Occasionally"	2 (Four-step Likert scale point 2)
	"Rarely"	1 (Four-step Likert scale point 1)
	Virtual middle	"I would say somewhere in the middle of this scheme."
	No response	Participant does not format his judgment.
	Response difficulties	I cannot decide which one.
	Criticizes response scheme	"These four categories are too few."
Mixes response format	Participant chooses a category which does not fit his judgment, and it is obvious that he mixes up the response format.	

matched in verbal protocol, and category III (*assessment not possible*) was used when information in the verbal protocol was too little to evaluate whether the item criteria were met or not.

The reliability of this procedure was assured by an independent rater who had substantial mindfulness knowledge. Rater concordance was measured using Cohen's kappa (see Wirtz and

Table 4 Examples of the qualitative content analysis procedure for one segment of a verbal protocol for one participant of MEG and of CG, respectively

Group	Comprehension segment	Paraphrasing	Summary
MEG	"... yes, this is exactly what is practiced in vipassana ... to watch this stream of thoughts and through concentration always and always again on the breath or a prickling in the body or on a feeling, that is to come back always to this and to me..."	That is the exercise which vipassana is all about, perception of cognition, emotions, breath, body sensations, and the conscious relocation of myself and my experiences to the present moment.	Item describes vipassana exercise
CG	"I think it is very important to be open because this is what keeps us young and helps us to get along with our everyday troubles."	Openness helps people to stay young and makes them able to master difficulties of life.	"To be open" as a possibility to stay young and to manage problems

Both statements were taken out from verbal protocols of item 1: "I am open to the experience of the present moment"

Caspar 2002). In accordance with Greve et al. (1997), kappa values below 0.4 are low, from 0.4 to 0.6 are intermediate, and values over 0.75 are high.

Development of a Coding System The second analysis strategy consisted of an adaptation of Chi's (1997) verbal analysis method. A coding system was developed following the work of Bickart and Felcher (1996) and Brown and colleagues (2009). A coding system categorizes text segments of the participants regarding their respective cognitive processes. Chi outlines eight basic steps that are followed to a greater or lesser extent depending on the aims of the research in question. The following two steps were seen as relevant to the present study, segmenting the protocols and developing a coding scheme. Segmentation was realized by identifying separate thoughts, ideas, or cognitive processes within the protocols. These segments served as units of analysis for further categorizations. In a theoretical deductive manner, all segments of all verbal protocols were sorted into the major three of the five categories of Tourangeau's model of cognitive processes used to answer survey questions (Tourangeau 1984), i.e., *comprehension*, *judgment*, *response*. *Comprehension* means the perception and individual interpretation of an item including verbalized problems in comprehension or criticism of item formulation. *Judgment* as a cognitive process in survey research is defined as all verbalizations that are immediately relevant for the response. The extent to which a respondent references an item statement to him or herself—the degree of self-referencing—is particularly interesting in that regard, because that is what is intended by a questionnaire in general (see Turner and Fiske 1968). *Response* includes the formal chosen response category of the Likert scale and further relevant information such as criticism of the response scale. The coding scheme and examples of coded segments are shown in Table 3. In a next step, subcategories were generated. For example, the cognitive process comprehension was inductively subcategorized into non-comprehension and further into explicit and implicit non-comprehension. Explicit non-comprehension includes direct verbalizations of non-comprehension (e.g., “I don't understand the question”), whereas the category implicit non-comprehension includes verbalizations indicating that the person had difficulties in understanding the question (e.g., “How am I supposed to pay attention to what's behind my actions?” as a verbalization to item 5: I pay attention to what's behind my actions.).

The final coding schema consists of 3 main and 26 subcategories (see Table 3 for all categories). A general rule for the application of the coding scheme was the assignment of only one category per segment. Furthermore, a category could not be given more than twice to every segmented verbal protocol (i.e., all statements of one person for one item). If a participant's verbal protocol for one item was, for example, separated into five segments (cognitive processes), then only a maximum of two out of these five segments

could be coded with the same category even if this category would also apply to a third segment. This procedure was chosen in order not to overestimate the impact of redundant statements. The reliability of this approach was assured by comparing the coding of the main rater with the codings of a second and independent rater, measured using Cohen's kappa. Both raters were unaware of the rating of the respective other rater (blinded rating).

Qualitative Content Analysis A qualitative analytic approach, based on the qualitative content analysis of Mayring (1983, 2002), was adopted for gathering information about specific comprehension patterns of the FMI by the participants of both groups.

First, all verbal segments of the comprehension category *defining item* from the above coding system was chosen. This category contains all segments of the verbal protocols where participants made an attempt to define the item for themselves (see Table 3: defining item for an example). Also, all statements of the original non-segmented transcribed verbal protocols which contained implicit hints of item comprehension were taken into account. In a second step, the chosen segments of the verbal protocols were paraphrased and summarized to their main content concerning an underlying proposed comprehension concept of the item or a part of the item (see Table 4 as an illustration). Afterwards, the summarized verbal protocol contents were organized and interpreted according to their structural similarities and differences within and between groups. The qualitative content analysis was conducted by the first author. Discussion among all authors, some of whom have a long-standing experience in qualitative research (especially the author Gabriele Lucius Hoene), was used in an attempt to validate the procedure.

Results

A total of 308 verbal protocols were obtained, 154 for each group. The number is based on 11 participants in two groups multiplied by 14 items. The groups did not differ in their ability to speak fluently (verbal fluency) or their processing time (see Table 5).

Table 5 Mean verbal fluency and processing time for both groups

	<i>n</i>	Verbal fluency M (SD)	Processing time M (SD)
MEG	11	115.4 (26.4)	807.7 (357.0)
CG	11	117.2 (24.7)	697.5 (199.9)

Verbal fluency and processing time in minutes; processing time for all items

Results of the External Item Criteria

Table 6 displays the results of the ratings of each of the 308 verbal protocols for each item comprehension in relation to the external item criteria. Out of them, 100 could not be classified (coded assessment not possible (III)) because of too little information in the statement of the participant. This category was even distributed among the two groups with 52 protocols in the MEG and 48 in the CG. Of the remaining 208 protocols, 97 were coded category I (mindfulness comprehension) and 111 category II (no item criteria fulfilled). If these two categories were split up for the two groups, the protocols of the MEG group were mainly assigned mindfulness comprehension (81 of 102, 79 %) and only rarely *no criteria fulfilled* (21 of 102, 21 %). For the comparison group, the situation was different. Here 16 of 106 protocols were assigned mindfulness comprehension (15 %) and 90 (85 %) no criteria fulfilled. The Chi-square test for the 2×2 table resulted in a highly significant contingency ($p < 0.001$, $\chi^2 = 86.4$, $df = 1$).

For the whole CG, the category mindfulness comprehension was assigned no more than twice to any single item, indicating that the intended criteria were almost never reflected in the mentation of the participants naïve to the concept of mindfulness. The CG achieved high values in category II for items 1, 2, 3, 4, 7, 8, 9, 11, and 14 (>50 % of group size). To the contrary, the MEG scored high in category I for the following items (>50 % of group size): items 1, 2, 3, 6, 7, 9, 10, and 14, indicating that the MEG had a good comprehension of mindfulness.

The interrater reliability between main and independent blind rater for all categories was $\kappa = 0.35$ and thus low. This low concordance was due to the fact that raters did not agree over category III, assessment not possible. As displayed above, approx. one third of verbal protocols contained too little relevant coding information and were thus difficult to code. However, if category III, assessment not possible, was omitted from the interrater reliability analysis, the kappa rose to an acceptable level of 0.78 for categories I and II, reflecting a high level of concordance.

Table 6 Results of the external item criteria procedure

	<i>n</i>	I Mindfulness comprehension	II No item criterion fulfilled	III Assessment not possible	Total
MEG	11	81	21	52	154
CG	11	16	90	48	154

Total consists of 22 participants (over both groups) multiplied by 14 items of the FMI=308 (154 for each group)

Results of the Coding Scheme

Comprehension *Comprehension problems* were coded with a ratio of 3:1 for the CG (27 CG/9 MEG). On a more detailed level, there were 15 *explicit* and 12 *implicit comprehension problems* for the CG and one explicit and eight implicit comprehension problems for the MEG. At item level, for the CG, comprehension problems were high (>50 % of group size) for items 3, 7, and 12. Item 8, which was three times coded implicit comprehension problems, was the most incomprehensible item for the MEG. The category *criticizes item* was coded for the CG 24 times (15 MEG, ratio 5:3), especially frequent for item 2 (four times) and item 6 (five times). The MEG especially was not satisfied with the wording of item 10 (three times). The groups did not differ largely in *defining item* (42 CG/56 MEG). The code *restates item* was coded approx. one third more often for the CG (90 CG/63 MEG) *considerations* with a ratio of about 3:1 (22 CG/7 MEG).

Judgment Over all items, both groups were similar in the degree of self-referencing (i.e., self-descriptive). At item level, items 1, 2, and 14 received the fewest and items 3, 6, and 12 the most self-descriptive verbalizations for the CG. The MEG showed no item with few ratings for the self-reference categories; the most self-reference descriptions were obtained for item 11 and 12. *Personal experience* in mindfulness was only coded for the MEG which was expected as it was an exclusion criterion for the CG. In this sense, the category served as a validation criterion for the independent variable. *Negative item assessment* examples were coded 13 times for CG (three for MEG; ratio, 4:1). Item 7, with three codings, was the most frequent single item for negative item assessment in the CG. Participants of the MEG verbalized *positive item assessments* (seven) more often, and this was only obtained once in the CG. The MEG received about the same amount of *agreement* as the comparison group but got much less *disagreement* (18 CG/6 MEG; ratio, 3:1).

If one merges the categories *self, descriptive* and *self, specifying, complex*, which are hard to differentiate, the CG got low ratings for items 1, 7, and 8 and high ratings for comprehension problems (item 7) and criticizes item (item 8).

Response The MEG shows a larger amount of affirmative categories than the comparison group and less negative responses. The mean sum score for the MEG is 40.36 (SD=4.08); for the CG, 37.10 (SD=4.41). The mean score of the CG is only marginally lower than the mean score of the MEG, a finding documented by previous research (Leigh et al. 2005). The answer category “almost always”

was chosen in the CG particularly often for items 1, 4, and 8 (>50 % of group size). Items 2, 3, and 12 were often declined (six, nine, and seven times). These items also obtained high ratings within the *comprehension difficulties* and *item critic* categories (six, eight, and six times).

Interrater Reliability Interrater reliability according to Cohen's kappa for the coding system over all categories was at an intermediate to high level (0.69). If the conceptually similar and therefore hard to differentiate categories self, descriptive and self, specifying, complex were merged, kappa was slightly higher (0.74). After consensual discussion between both raters regarding the coding of *comprehension* and *judgment* categories, reliability rose to 0.80. If the non-controversial response categories were omitted the kappa dropped to 0.63, still showing intermediate reliability. The results of the coding scheme procedure can be seen in Table 7.

Table 7 Results of the coding system procedure

Cognitive process	Content	MEG n=11	CG n=11
Comprehension	Explicit C. problems	1	15
	Implicit C. problems	8	12
	Criticizes item	15	24
	Defining item	56	42
	Consideration	7	22
	Restates item	63	90
	Pause	15	2
Judgment	Personal experience	19	0
	Self, descriptive	113	107
	Self, prescriptive	33	33
	General statement	39	32
	Specifying, simple	17	11
	Self, specifying, complex	52	53
	Agreement	22	25
	Disagreement	6	18
Response	Positive item assessment	7	1
	Negative item assessment	3	13
	“Almost always”	29	32
	“Fairly often”	81	43
	“Occasionally”	25	40
	“Rarely”	3	20
	Virtual middle	12	4
	No response	4	15
	Response difficulties	10	19
	Criticizes response scheme	5	12
Mixes response format	0	5	

Results of Qualitative Analysis

The results of qualitative content analysis demonstrate striking differences in semantic item comprehension between the groups. We present summaries of two verbal protocol analyses for items 3 and 7 here in detail for both groups. A brief summary of the main results according to the qualitative analysis for every item will also be given for the CG. The MEG comprehended items overall in the intended way, which for reasons of space is not presented here for every item. However, short comparative references to the MEG are made when seen as appropriate. The numbers in brackets represent participant ID.

Item 3: When I notice an absence of mind, I gently return to the experience of the here and now.

MEG: For the MEG, item 3 represented mindfulness meditation per se: “this is what I practice in mindfulness meditation” (3), “this describes exactly mindfulness meditation” (9, 14, 15). Some participants of the MEG emphasize the affirmative character of the item statement: “this is a nice instruction for a guided meditation” (2). Another participant differentiates the application of the item to formal meditation and everyday life: “This is something which I can do better in meditation than in everyday life” (14). Furthermore, the term “gentle” in the item statement is recognized as an accurate description of the “return to the experience of the here and now”: “... and the word gentle especially is very helpful” (15), “this feels like a gentle landing in the present moment” (12). For the MEG, a “gentle return to the experience of the here and now” means aligning oneself again to the actual moment and to the observation of thoughts, feelings, breath, and body sensations (3, 4), “without judging oneself” (2) respectively “an overall caring handling with oneself” (15, 16).

CG: The CG comprehend this item statement in a totally different way as an absence of thought (dissociation) from an actual everyday activity like working: “...this means that I had to do something and let my mind wander” (12), “... absence of thought from something I had to do” (1, 5, 7, 10, 11). The “return to the experience of the here and now” is recognized as a sudden waking up and simply continuing of work: “this is like a flip which flashes through my mind and then I continue work where I have stopped” (2, 6, 8). The item term “gentle” seems to be an irritating formulation, because the reason for a “gentle return” is not familiar to CG participants or is regarded as an “esoteric” (11) formulation of an everyday phenomenon.

The MEG participants comprehended item 3 totally different from the CG. No participant of the MEG comprehended this item as an everyday phenomenon—as being absent and then returning suddenly to a certain activity. In most cases, for the MEG, this item accurately represented formal mindfulness practice. As with the comparison group, absence was interpreted as an absence of thought—this is a similarity between the groups. This finding is especially important for the German version of the FMI, where the term “mind” is not included in the item wording. The word “gentle,” which confuses most participants of the CG, meets with a positive response in the MEG.

Item 7: I feel connected to my experience in the here and now.

MEG: The MEG interpreted the term “to be connected with my experience” as an awareness of present feelings, thoughts, body sensations: “for example, when riding a bicycle, at the same time I notice the wind” (16), “this means noticing my body sensations, my emotions, and my cognitive experiences” (14), “this means that I am aware of the present situation as I experience it in this very moment rather than being absorbed by thoughts of the future or the past” (9).

CG: The CG on the other hand understood the term “experience” in the sense of a personal, biographical experience of life in general: “sure, my experience made me who I am; without my experience, I would not be the person I am” (12, also 10, 11, 7, 5, 1). The formulation “to be connected with” is irritating for most participants of the CG: “this sounds esoteric” (2), “otherwise, I would not be self-determined” (5), or even not comprehensible: “this makes no sense to me” (11). Some participants of the CG differentiate between a conscious and an unconscious way of being in contact: “unconsciously, I am definitely always connected with my experience, definitely not consciously” (12, 7), and being conscious in “contact with their experience” especially in situations in which a decision is to be made: “consciously, I get in contact with my experience in decision-making situations—there I draw on a wealth of experience.” (12, 8, 7) or “I try to become conscious of my (biographical) experiences—to know which one has led me to a decision” (7). The item add on “here and now” is mixed up with the actual interviewing situation: “now I am not connecting with my experience of life because I am concentrating on this task right now” (6).

In conclusion, items 3 and 7 are examples of how some items of the FMI are misinterpreted by participants without meditation experience. The term “experience” in particular is frequently misunderstood as a biographical past experience of life. These conclusions are backed up by the qualitative results from other items:

Item 1: I am open to the experience of the present moment.

The CG interpreted the item in the sense of a question about the personality attitude “openness.” The term “experience” was often misinterpreted in the sense of (past) experience of life and was irritating for some participants. Some participants of the MEG stressed the affirmation character of the item and said that this is something they say to themselves.

Item 2: I sense my body, whether eating, cooking, cleaning, or talking.

The CG read this item as the ability to notice bodily sensations such as pain or illness or as an overall question of body reference. The term “I sense my body” (in its German translation) sounded strange for some participants or was completely unknown. Simultaneous “sensing” while performing the mentioned activities was noticed as difficult, impossible, or absurd.

Item 4: I am able to appreciate myself.

The item was well comprehended by the CG. The CG's verbalizations, however, were stated in the context of personal performance and social life, whereas the MEG verbalized the context of their meditation practice and an overall compassionate attitude towards the self as a source for self-appreciation.

Item 5: I pay attention to what's behind my actions.

The CG comprehended item 5 as meaning a retrospective analysis of their own actions in the sense of a prospective consideration of consequences of actions for important decision-making situations or even in a moral sense. The MEG, on the other hand, interpreted the item in the sense of a “continuous action monitoring” or a mere registration of impulses, emotions, behaviors of the self in the present moment, respectively.

Item 6: I see my mistakes and difficulties without judging them.

In this item, two statements are merged, and this created some confusion in the CG. The first and second parts were answered differently by some participants (1, to see one's mistakes, vs. 2, to not judge them). The combination of the statements also influenced the responses. Some people responded to the item with the answer category *rarely* according to the interpretation: “Do you

judge yourself for mistakes ...?"; others choose almost always which, as one can see from the verbalizations, was intended to refer to observing mistakes while not judging them. The word "judging" was evaluated by the CG as a very strong formulation compared to the MEG.

Item 8: I accept unpleasant experiences.

This item was seen by the CG as a statement regarding past experiences of life, whereas the MEG stressed the present-moment aspect.

Item 9: I am friendly to myself when things go wrong.

Participants of the CG often asked themselves what was meant by "when things go wrong." The term "to be friendly with oneself" was interpreted in terms of "to be able to forgive oneself." Participants of the MEG, on the other hand, stressed that the expression "when things go wrong" entailed a judgment which they were able to abstain from.

Item 10: I watch my feelings without getting lost in them.

For the CG, this item consisted of two statements that had been merged into one (as per item 6). "Getting lost in them" was evaluated as a very strong formulation by the CG, more appropriate would be *without being worked up by them*. "To watch feelings" was read in the sense of a cognitive reflection on emotions, similar to "I analyze my feelings, without getting caught up by them." Overall, the item was understood by the CG as "I am an emotional person."

Item 11: In difficult situations, I can pause without immediately reacting.

The CG pointed out that the formulation "I can pause" is ambiguous, because one may be in principle able to make a break but still does not do so. For most participants in the CG, "I can pause" meant to think about behavior strategies to solve the "difficult situation," whereas the MEG mentioned the conscious setting up of a distance from "the difficult situation" in order to "defuse" the difficult situation automatically.

Item 12: I experience moments of inner peace and ease, even when things get hectic and stressful.

The wording of the item in the English FMI version is too different from the German version to convey the results of the qualitative analysis. A more literal translation from the original would be "...even when external pain and restlessness are present." The words "pain" and "external," especially, caused most of the difficulties for the CG in the German version.

Item 13: I am impatient with myself and with others.

Typical for this item (in both groups) was a splitting of the response into the two statements entailed

in this item (to be impatient with myself/to be impatient with others). Both groups were similar in their verbalizations. The item was well comprehended by the CG.

Item 14: I am able to smile when I notice how I sometimes make life difficult.

For the CG, "to smile" in this item meant something like a smile of relief after the experience of an avoidable or needless "making life difficult" situation. "To make life difficult" was associated by the CG with being too demanding with the self or setting oneself too many rules. For the MEG, "smiling" seemed to imply creating a certain distance from oneself in order to cope with the difficult aspect.

Overall, the MEG mainly comprehended items in the intended way. In contrast, only items 4, 6, 9, 11, 13, and 14 could be interpreted as being adequately comprehended by the CG in the intended way according to the qualitative item analysis. Only these items can be, despite differences, considered as generally understandable for participants without mindfulness experience.

Discussion

We conducted a mixed method but mainly qualitative study with three different approaches to assess the item comprehension of the FMI in meditators and non-meditators. Within qualitative social research, different quality criteria for the validity of results are recommended, all of which were included in the present study: (1) a triangulation of methods through different methods from different perspectives (Denzin 1978), (2) transparency of the research process, and (3) a consensual validation with mindfulness and methodology experts (see also Flick 1992).

Results of the present study indicate that many items of the FMI Short Version are difficult or impossible to understand for persons without mindfulness experience. This main result is congruent with Grossman's (2008) assumption that the semantic interpretation of questionnaire items depends on meditation/mindfulness experience. This is especially true for items 1, 2, 3, 5, 7, 8, 10, and 12 almost exactly. A closer look at these items reveals that they all semantically represent the awareness/present-moment facet of mindfulness according to the two-factor concept awareness/acceptance of mindfulness of Bishop et al. (2004). Interestingly, these results conform to recent findings arguing that mindfulness should be conceived as a two-factorial construct consisting of an awareness and an acceptance factor (Sauer et al. 2011). From a methodological point of view, it may be stated that

qualitative or mixed-method approaches, such as the present study, may be an effective way of fostering the understanding of the construct of mindfulness.

Thus, our data show that it is necessary to be at least intellectually informed about the concept of mindfulness in order to understand this group of items properly. Whether a deeper experience of mindfulness is necessary to comprehend the items is not possible to decide on the basis of our data and needs to be tested in separate study. On the other hand, the items of the FMI which semantically represent the concept of a non-judgmental or accepting attitude were much better understood by non-meditators.

Another important result of the present study is the fact that almost all of the mindfulness-experienced participants comprehended the items of the FMI in the intended way in a relatively homogeneous fashion. This indicates that the items represent the core of the “mindfulness concept” very well, although the meditation group varies in the extent of meditation experience from 4 to 40 years. This finding is contradictory to Grossman's (2008) assumption that item interpretation depends on the extent of meditation experience.

There are a number of limitations of the present study that should be mentioned here: The method to assess item comprehension by external criteria (procedure 1) was especially designed for our study and was applied for the first time to such a context. Here it revealed that based on the think aloud method, not all verbal protocols were of sufficient length and detail to be judged. Approx. one third of the protocols had to be discarded, and the analysis could only be conducted on the remaining ones. Also the interrater reliability was low for all protocols and rose only to an acceptable $\kappa=0.78$ once the protocols with too little information were discarded. Moreover, it may have been the case that the occurrence of linguistic markers for mindfulness-related language used by the MEG had an influence on raters. That means raters may have overestimated the suitability of criteria for judging item comprehension by the MEG. However, the MEG group obtained low scores in items 4, 5, and 12, indicating a certain sensitivity to the external item criteria. Based on the results of procedure 1, it can be concluded that the FMI is not fully understandable for non-meditators (only 15 % met the criteria), while the mindfulness-experienced group on the other hand shows an acceptable understanding (85 % met the criteria). This is the case for the results of the main rater as well as for the ratings of the independent rater supporting the main result.

Regarding the results of the coding scheme method (procedure 2), it was revealed that there was a higher number of implicit and explicit comprehension difficulties and criticism from the CG than from the MEG, especially for items 2, 3, and 12. It is worth noting that exactly the same items received the highest amount of negative responses from the CG. Concerning cognitive processes of judgment, the

groups did not differ significantly. This is surprising since we expected the CG participants to show decreased verbalizations of self-reference due to their lack of familiarity with the concept and thus no personal experiences in mindfulness practice. The opposite is true. Regardless of whether the person understood the item correctly or not, a reason as to why they were being asked was constructed. An explanation for this result could also be the so-called “non-attitudes” research (Converse 1970). Sometimes people give answers to questions regarding a certain attitude which is definitely not true for them (Schuman and Presser 1981). This indicates that respondents generally act on the fundamental assumption that a question has something to do with them just because they are asked this very question. Thus, the verbalized degree of self-reference is probably not a satisfactory predictor of the adequacy of judgment—as a basis for response. This interpretation is somewhat contradictory to Turner and Fiske (1968), who concluded that the adequacy of the response to a questionnaire depends on the verbalized degree of self-reference while responding. In other words, although the CG participants did not understand most of the items in the intended way, they still constructed some alternative meaning which in turn resulted in high rates of self-referencing and thus in a high sum score of the overall responses. This interpretation would explain why, in many samples of mindfulness-naïve participants, high scores of self-attributed mindfulness can be found. This result clearly points out that, at least for the FMI, these scores in naïve persons cannot be solely attributed to the concept of mindfulness itself. This conclusion is further supported by the satisfactory kappa scores for the interrater reliability of the coding scheme. It should be critically remarked that *retrieval* and *judgment* processes could not be differentiated by means of our coding system. Future research should work out indications for a differentiation between *retrieval* and *judgment* processes as well as specific rules for the categorization of these two cognitive processes.

Finally, the qualitative item analysis (procedure 3) proved to be insightful. It revealed that items containing the present moment or awareness aspect of mindfulness are much less understandable than items regarding the acceptance aspect for the CG. It could also be illustrated that the comprehension in the MEG was highly homogenous. Their familiarity with the mindfulness language expressed itself in a pronounced flexibility when applying specific terms. For example, they anticipated aspects of mindfulness being handled later in a following item or vice versa. This did not occur in the CG, who had great comprehension problems especially with items containing very “technical” mindfulness terms. They understood them in an everyday meaning which can be profoundly misleading (e.g., for the term “experience”, e.g., Item 7). Furthermore, it demonstrated the tendency of non-meditators to construct the meaning

of the items in a context of personal retrospective decision-making situations and personal biographical experience of life in general, especially in items 5, 7, and 11. Also, a confusing or ambiguous wording of some items was criticized by these participants, e.g., as being too ambiguous or blurry. Reasons for this criticism could be worked out clearly by means of the qualitative approach. The qualitative item analysis also uncovered the specific reasons for the comprehension difficulties of many items. The partly strong negative item assessments found when coding statements for this group were revealed by the qualitative analysis as being basic misinterpretations of item terms and not a pure general or arbitrary denial of the item statement. In summary, the qualitative analysis proved to be complementary to the results of the other methods used.

Regarding future research, we would like to stress that the problems mentioned here may not be specific to the FMI, but may also concern similar mindfulness inventories using a self-report format. Whether this is true or not can be found by conducting similar assessments of item comprehension for other scales. The overarching question here is whether it is possible to formulate mindfulness in a language which does not presuppose knowledge about the concept mindfulness. Or is the state or trait of being mindful dependent on the language in which it is expressed? For future studies, it is necessary to examine the interface between the specific language of mindfulness and everyday language in order to find an answer to these questions. Card Sorting Techniques (Prüfer and Rexroth 2005) provide a possible research strategy which could be used to allow mindfulness experts to rate terms for their suitability to capture mindfulness and for non-meditators to rate those terms for comprehension. Furthermore, research on what defines *mindfulness behavior* should be formulated in an effort to judge persons regarding their actions, rather than regarding their knowledge of the mindfulness concept.

We conclude that a “bare” “awareness of the present moment,” which includes vigilant attention and an attitude of non-judgment, is obviously less rooted in our western culture. Thus, it can be challenged whether questionnaire items assessing this concept have sufficient construct validity without some kind of knowledge and/or experience of mindfulness training. Recently, some research on so-called “dispositional mindfulness” was published (see e.g., Creswell et al. 2007, p. 560–65; Way et al. 2010, p. 12–24), in which a sample of participants with no mindfulness experience was split on the basis of questionnaire results into high or low scorers in order to draw further conclusions. Our results suggest that this notion of “dispositional mindfulness” may due to a different comprehension of the questionnaire items by mindfulness-naïve participants and does not refer exactly to what is usually understood as the concept of mindfulness. This issue should be clarified in future

studies to see whether the idea that a natural variation of mindfulness can be assessed with self-report instruments in non-meditators is valid.

Whether mindfulness questionnaires other than FMI suffer from similar drawbacks remains to be answered by future studies. However, given the similar structures of many mindfulness scales, one can assume that this problem may be also present in similar scales. If questionnaires on the self-attribution of mindfulness have a limited construct validity (especially for items tapping into the awareness facet of mindfulness) in non-meditators, then many results of studies published so far have to be challenged. In this case, the use of our scale and maybe also of others to control for the intended effect of a mindfulness intervention, e.g., by administering a scale before and after a MBSR course in a clinical sample which begins the study with no experience of meditation, might lead to biased results. This is due to the fact that, semantically, items are easily misinterpreted and elicit different self-referential processes before the intervention and thus cannot be compared to the post-intervention data. In contrast to the assessment of many other psychological constructs, the assessment of self-attributed mindfulness in self-report scales has a much more intimate relationship to the concept itself. This is due to the fact that some scales draw explicit on the conceptual knowledge taught in the intervention rather than making an indirect assessment by inferring from statements with no explicit mentioning of mindfulness key concepts. This may also apply for studies which correlate the change in self-attributed mindfulness to other variables such as neurophysiological indicators or in moderator/mediator analyses. The only suitable use of these widely applied scales might be the assessment of samples which have some prior mindfulness experience. Our results indicate that for such a sample, comprehension and thus construct validity can be achieved. Then, and only then, the mindfulness questionnaire may indeed measure what it was intended to measure.

Acknowledgments We would like to thank our participants who have given their time to take part in this study. We would like to thank Alf Martin and Andreas Schmidt for helping with the rating, Harald Walach for providing expert criteria. In addition, we like to thank Kathrin Simshauser, Christine Kalweit and Attila Goerog for proof-reading the manuscript.

References

- Baer, R. A. (2007). Mindfulness, assessment, and transdiagnostic processes. *Psychological Inquiry*, 18(4), 238–242.
- Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: the Kentucky Inventory of Mindfulness Skills. *Assessment*, 11, 191–206.

- Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*(1), 27–45.
- Bickart, B., & Felcher, E. M. (1996). Expanding and enhancing the use of verbal protocols in survey research. In N. Schwarz & S. Sudman (Eds.), *Answering questions: methodology for determining cognitive and communicative processes in survey research* (pp. 115–142). San Francisco: Jossey-Bass.
- Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., et al. (2004). Mindfulness: a proposed operational definition. *Clinical Psychology Science and Practice, 11*, 230–241.
- Brown, K. W., & Ryan, R. M. (2003). The benefit of being present: mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848.
- Brown, G., Hawkes, N. C., & Tata, P. (2009). Construct validity and vulnerability to anxiety: a cognitive interviewing study of the revised Anxiety Sensitivity Index. *Journal of Anxiety Disorders, 7*(23), 942–949.
- Cardaciotto, L., Herbert, J. D., Forman, E. M., Moitra, E., & Farrow, V. (2008). The assessment of present-moment awareness and acceptance: the Philadelphia Mindfulness Scale. *Assessment, 15*(2), 204–223.
- Chadwick, P., Hember, M., Symes, J., Peters, E., Kuipers, E., & Dagnan, D. (2008). Responding mindfully to unpleasant thoughts and images: reliability and validity of the Southampton Mindfulness Questionnaire (SMQ). *British Journal of Clinical Psychology, 47*, 451–455.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315.
- Converse, P. E. (1970). Attitudes and non-attitudes: continuation of a dialogue. In E. Tufté (Ed.), *The quantitative analysis of social problems* (pp. 168–189). Reading: Addison-Wesley.
- Creswell, D., Way, B. M., Eisenberger, N. I., & Lieberman, M. D. (2007). Neural correlates of dispositional mindfulness during affect labeling. *Psychosomatic Medicine, 69*, 560–565.
- Denzin, N. K. (1978). *The research act*. New York: McGraw-Hill.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: verbal reports as data* (Revth ed.). Cambridge: Bradford books/MIT Press.
- Fattah, S. (2009). *Kommunikative Konstruktion von Achtsamkeit in Erzählinterviews: Qualitative Studie mit Teilnehmern eines MBSR-Kurses*. Unpublished thesis, Institute of Psychology, Albert-Ludwigs-University Freiburg, Germany.
- Feldman, G. C., Hayes, A. M., Kumar, S., Greeson, J., & Laurenceau, J.-P. (2007). Mindfulness and emotion regulation: the development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). *Journal of Psychopathology and Behavioral Assessment, 29*, 177–190.
- Flick, U. (1992). Triangulation revisited: strategy of validation or alternative? *Journal for the Theory of Social Behaviour, 22*(2), 175–197.
- Greve, W., Wentura, D., Gräser, H., & Schmitz, U. (1997). *Wissenschaftliche Beobachtung: Eine Einführung* (2nd ed.). Weinheim: Beltz.
- Grossman, P. (2008). On measuring mindfulness in psychosomatic and psychological research. *Journal of Psychosomatic Research, 64* (4), 405–408.
- Güthlin, C. (2004). Response shift: Alte Probleme der Veränderungsmessung, neu angewendet auf gesundheitsbezogene Lebensqualität. *Zeitschrift für Medizinische Psychologie, 13*, 165–174.
- Haigh, E. A. P., Moore, M. T., Kashdan, T. B., & Fresco, D. M. (2011). Examination of the factor structure and concurrent validity of the Langer mindfulness/mindlessness scale. *Assessment, 18*, 11–26.
- Hak, T., van der Veer, K., & Jansen, H. (2008). The Three-Step Test Interview (TSTI): an observation based method for pretesting self-completion questionnaires. *Survey Research Methods, 2*(3), 143–150.
- Huber, G. L., & Mandl, H. (Eds.). (1982). *Verbale Daten: Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung*. Weinheim: Beltz.
- Kuckartz, U. (2006). Quick and dirty? Qualitative Methoden in der drittmittelfinanzierten Evaluation in der Umweltforschung. In U. Flick (Ed.), *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen* (pp. 267–283). Reinbeck: Rowohlt.
- Lau, M. A., Bishop, S. R., Segal, Z. V., Buis, T., Anderson, N. D., Carlson, L., et al. (2006). The Toronto Mindfulness Scale: development and validation. *Journal of Clinical Psychology, 62*(12), 1445–1467.
- Leigh, J., Bowen, S., & Marlatt, G. A. (2005). Spirituality, mindfulness and substance abuse. *Addictive Behaviors, 30*(7), 1335–1341.
- Marcel, A. J. (2003). Introspective report: trust, self-knowledge and science. *Journal of Consciousness Studies, 10*(167), 186.
- Mayring, P. (1983). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz: Weinheim.
- Mayring, P. (2002). *Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken*. Weinheim: Beltz Studium.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., et al. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly, 68*(1), 109–130.
- Prüfer, P., & Rexroth, M. (2005). *Kognitive Interviews*. Mannheim: ZUMA How-to-Reihe Nr. 15.
- Reavley, N., & Pallant, J. F. (2009). Development of a scale to assess the meditation experience. *Personality and Individual Differences, 47*, 547–552.
- Sauer, S., Walach, H., Schmidt, S., Hinterberger, T., Horan, M., & Kohls, N. (2011). Implicit and explicit emotional behavior and mindfulness. *Consciousness and Cognition, 20*, 1558–1569.
- Schneider, J. F., & Reichl, C. (2006). Exploring ease in thinking aloud. *Psychological Reports, 98*(1), 85–90.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: experiments on question form, wording, and context. quantitative studies in social relations*. New York: Academic.
- Solloway, S. G., & Fisher, W. P. (2007). Mindfulness practice: a Rasch variable construct innovation. *Journal of Applied Measurement, 8* (4), 359–372.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: a practical guide to questionnaire design (1 ed.)*. The Jossey-Bass series in social and behavioral science series. San Francisco: Jossey-Bass.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: building a bridge between disciplines* (pp. 73–100). Washington, DC: National Academy Press.
- Turner, C. B., & Fiske, D. W. (1968). Item quality and appropriateness of response processes. *Educational and Psychological Measurement, 28*(2), 297–315.
- Walach, H., Buchheld, N., Buttenmüller, V., Kleinknecht, N., & Schmidt, S. (2006). Measuring mindfulness—the Freiburg Mindfulness Inventory (FMI). *Personality and Individual Differences, 40*, 1543–1555.
- Way, B. M., Creswell, J. D., Eisenberger, N. I., & Lieberman, M. D. (2010). Dispositional mindfulness and depressive symptomatology: correlations with limbic and self-referential neural activity during rest. *Emotion, 10*, 12–24.
- Willis, G. B. (2005). *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks: Sage.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.