How You Measure Is What You Get: Differences in Self- and External Ratings of Emotional Experiences in Home Dreams

PILLERIIN SIKKA
University of Turku and University of Skövde

DIANA FEILHAUER
University of Skövde

KATJA VALLI and ANTTI REVONSUO
University of Turku and University of Skövde

This study demonstrates that different methods for measuring emotional experiences in dreams—self-ratings of dreams using emotion rating scales versus external ratings in the form of content analysis of narrative dream reports—can lead to strikingly different results and contradicting conclusions about the emotional content of home dreams. During 3 consecutive weeks, every morning upon awakening, 44 participants (16 men, 28 women, average age 26.9 ± 5.1 years) reported their dreams and rated their emotional experiences in those dreams using the modified Differential Emotions Scale. Two external judges rated emotional experiences in the same 552 ($M = 12.55 \pm 5.72$) home dream reports using the same scale. Comparison of the 2 methods showed that with self-ratings dreams were rated as more emotional and more positive than with external ratings. Moreover, whereas with self-ratings the majority of dreams was rated as positively valenced, with external ratings the majority of dream reports was rated as negatively valenced. Although self- and external ratings converge, at least partially, in the measurement of negative emotional experiences, they diverge greatly in the measurement of positive emotional experiences. On one hand, this discrepancy may result from different biases inherent in the 2 measurement methods highlighting the need to develop better methods for measuring emotional experiences. On the other hand, self- and external ratings may capture different phenomena and should thus be considered complementary and used concurrently. Nevertheless, results suggest that negative emotional experiences can be measured in a more valid and reliable manner than positive emotional experiences. Supplementary results are available at http://www.press.uillinois.edu/journals/ajp/media/sikka/how_you_measure/

KEYWORDS: emotions, dreaming, dream emotions, content analysis, gender differences

In emotion research, as in consciousness research in general, one of the most central and difficult issues is the measurement of subjective experience (Nielsen & Kaszniak, 2007). Emotional experience is an important phenomenal quality or feature of subjective experience. Being subjective by definition, there are no known objective external or third-person methods for measuring emotional experience (even though biological indicators and facial or vocal expressions can be used to infer certain subjective states) (Scherer, 2005). Thus, we have to rely on self-reports provided by individuals because these represent "the most reliable and possibly only window that researchers have on conscious, subjective, emotional experience" (Barrett, 1996, p. 47).

However, there are numerous conceptual and methodological challenges in using self-reports to measure emotional experience. A fundamental issue is whether and to what extent self-report reflects the actual "raw" emotional experience (phenomenal consciousness), as opposed to cognitive operations performed on these experiences, allowing them to be described and reported (reflective consciousness). By definition, self-report depends on reflective consciousness, and therefore we cannot be sure how much the report of the experience is skewed or altered in the process of reporting (for a detailed discussion on this issue, see Nielsen & Kaszniak, 2007, and Revonsuo, 2006). Other issues are related to the collection of self-reports, such as which type of self-reports are collected (e.g., using rating scales or narrative descriptions in a free-response format), which subjective states are measured (e.g., emotions or moods), how many and what type of emotions are measured (e.g., discrete or dimensions of emotions), whether the frequency or intensity of emotions is measured, what is the timing of the measurement (e.g., while emotions are experienced or retrospective reports), and who rates the emotions (e.g., participants themselves or researchers who analyze the content of the self-report) (see Larsen & Fredrickson, 1999, for a discussion on the various measurement issues in emotion research).

Different choices regarding the methodological issues described above may lead to inconsistent results and conclusions about the frequency and nature of emotional experiences. For example, one of the most well-known and frequently used standardized self-rating scales, the Positive and Negative Affect Schedule (Watson & Clark, 1994; Watson, Clark, & Tellegen, 1988), asks participants to rate on a 5-point scale the extent to which they have experienced a list of different "feelings or emotions" during a specified time period. Watson (2000) used this scale to assess students' momentary emotional experience once or several times a day. In more than 90% of the observations the students reported at least a moderate amount of emotional experiences, leading the author to conclude that "waking consciousness is experienced as a continuous stream of affect . . . such that people are always experiencing some type of mood" (Watson, 2000, p. 13). In contrast, Heavey and Hurlburt (2008) used an open-ended free response method, the Descriptive Experience Sampling (DES), to study students' ongoing "inner experiences" in natural environment at specific moments in time. In DES, participants are given a beeper, signaled randomly several times a day, and upon signaling asked to immediately recall and take notes of the details of their experience. Within 24 hr the participant is interviewed about the experience at each sampled moment, based on which the researchers code the experience using a codebook. Using this method, the authors found that respondents had emotional experiences in only 26% of the sampled moments, based on which they concluded that "feelings . . . [were] present only about a quarter of the time" (Heavey & Hurlburt, 2008, p. 11). These discrepant findings may result from measuring different subjective states, obtaining different types of self-reports, differences in who rates emotional experiences, or, as suggested by Heavey, Hurlburt, and Lefforge (2012), differences in the temporal precision of the measurement. Thus, the results and conclusions about emotional experiences depend on what is measured and how.

The same methodological challenges are faced in dream research when studying emotional experiences in dreams. A dream is a sequence of subjective conscious experience occurring during sleep that is recalled upon awakening (Pagel et al., 2001). As this definition implies, we do not have direct access to dreams (thus far there is no "dream catcher" device that would "read out" the contents of dreams online as they are happening; Revonsuo, 2006) but have to rely on (verbal or written) self-reports provided by the person after the experience. Therefore, retrospective

self-report measures are (at least at present) the only way to study dream experiences, including emotional experiences in dreams (Domhoff, 2003; Stickgold, 2017; Windt, 2013, 2015).

Studies investigating the emotional content of dreams vary widely with regard to the method for collecting and analyzing self-reports (for a detailed review of the various methods and their advantages and disadvantages, see Kahan & Horton, 2012; Schredl, 2010a; Windt, 2015; Zadra & Domhoff, 2017). Results differ depending on the particular combination of methods used, such as whether narrative dream reports are content analyzed by external judges or dream experiences are rated by participants themselves using emotion rating scales.

Collecting narrative dream reports and analyzing their content has been the standard method applied in dream research (Domhoff, 2003; Hall & Van de Castle, 1966; Winget & Kramer, 1979). With this method, participants are asked to report (either verbally or in a written form) their dreams upon awakening in as much detail as they can remember. Then external judges identify and code all emotions expressed in the dream report using a specific content analysis system or rating scale. Alternatively, external judges are sometimes instructed to rate not only explicitly expressed emotions but also implicit emotional experiences. For this, global rating scales measuring the overall emotional tone or intensity of the dream report can be used (see Schredl, 2010b for a review on the basic principles of dream content analysis). However, it has been questioned whether external judges are able to accurately and reliably infer emotional experiences from the dream content (Domhoff, 2005), and therefore caution against inferring anything beyond what is explicitly expressed in dream reports has been advised (Windt, 2015). Using content analysis, dream reports have been found to be mostly nonemotional (e.g., Snyder, 1970), with less than one emotion per dream report (e.g., Hall & Van de Castle, 1966), and to contain more negative than positive emotions (e.g., Hall, 1951; Hall & Van de Castle, 1966; Hsu & Yu, 2016; Snyder, 1970). These findings have led to the conclusion that dreams are predominantly negatively valenced, an assumption that underlies several well-known theories of dream function (e.g., Cartwright, 1996; Hartmann, 1995; Kramer, 1991; Revonsuo, 2000).

Instead of, or in addition to, collecting narrative dream reports, in other studies (e.g., Fosse, Stickgold, & Hobson, 2001; Kahn, Pace-Schott, & Hobson, 2002; Merritt, Stickgold, Pace-Schott, Williams, & Hobson, 1994) participants have been provided with affirmative probes that ask them to report specific information about dream emotions that may not be spontaneously reported. In such studies participants are asked to rate, upon awakening, the occurrence or intensity of emotions experienced in the dream using standardized or nonstandardized rating scales or checklists of emotions. With this method, most dreams have been rated to be emotional (e.g., Fosse et al., 2001; Hall, 1951; St-Onge et al., 2005; Strauch & Meier, 1996), with more than two emotions per dream (e.g., Blick & Howe, 1984; Merritt et al., 1994; Nielsen, Deslauriers, & Baylor, 1991). Results regarding emotional valence are varied, ranging from findings that dreams contain more negative than positive emotions (Blick & Howe, 1984; Merritt et al., 1994; Kahan & Claudatos, 2016) to those that show a balanced proportion of positive and negative emotions (Fosse et al., 2001; Yu, 2007) or emotional tone (Blagrove, Farmer, & Williams, 2004) to still others that show more positively than negatively valenced dreams and emotions in dreams (St-Onge et al., 2005). These diverse findings are probably due to other methodological differences between studies, such as data collection environment and procedure (e.g., St-Onge et al., 2005), the number and type of specific emotions measured (e.g., Kahan & Claudatos, 2016), and sample characteristics, such as age (e.g., Blick & Howe, 1984; St-Onge et al., 2005). Regardless, studies using self-ratings challenge the view that dreams are mostly negatively valenced (Sikka, Valli, Virta, & Revonsuo, 2014).

To explain such discrepant findings obtained with the two methods—external ratings of narrative dream reports and self-ratings using emotion rating scales—a systematic comparison of methods is needed.

Direct Comparisons of Self- and External Ratings of Dream Emotions

To date only two studies have been published (Schredl & Doll, 1998; Sikka et al., 2014) that have directly compared external ratings (ER) of emotions in narrative dream reports with self-ratings (SR) of

emotional experiences in dreams using emotion rating scales in the same participants. Schredl and Doll (1998) applied three different measurements of dream emotions to the same 133 home dream reports: SR using two global rating scales for measuring the intensity of the positive and negative emotional tone of the dream as a whole, ER in the form of content analysis identifying specific emotions in a dream with the Emotions scale of the Hall and Van de Castle (1966) content analysis system, and ER using the same two global rating scales as for SR. With SR the majority of dreams (99.2%), but with content analysis less than half of the dreams (42.1%), were rated as emotional. When external judges used the global rating scale, more than two thirds of the dreams (86.5%) were rated as emotional. Regarding emotional valence, the study showed that the intensity of both positive and negative emotions was rated higher with SR, as compared to ER, although the difference was larger for positive than for negative emotions. When the two rating methods were looked at separately, with ER there were more than twice as many negative than positive dreams, regardless of whether content analysis (26.3% and 9.0%, respectively) or global rating scale (56.4% and 21.1%, respectively) was used. With SR the proportion of negative and positive dreams was more balanced (50.4% and 36.8%, respectively), although negatively valenced dreams still prevailed. No gender differences in dream emotionality were found with either measure.

Sikka et al. (2014) measured the occurrence of specific emotions in 115 laboratory dream reports using the same scale (modified Differential Emotions Scale; Fredrickson, 2013) for both SR and ER. Participants rated the intensity of emotions they experienced in the dream (although for comparison with ER, the analyses were based on the occurrence of emotions), and external raters identified and coded all emotions in the dream report that were either explicitly expressed or could be unambiguously inferred from the behavior of the dream self. Whereas with SR all of the dreams (100%), with ER only about one third of the dream reports (28.7%), were rated as emotional. Moreover, SR yielded not only a larger number of emotional dreams but also a larger number of emotions per dream. Regarding emotional valence, with SR a larger number of dreams were rated to be

positive and to contain more positive and negative emotions than with ER, although the difference was larger for positive than for negative emotions. When the two rating methods were looked at separately, SR dreams were rated to be more positive than negative (79.1% vs. 12.2%), whereas with ER the proportion of dream reports rated to be positive and negative was rather balanced (9.6% vs. 11.3%). The authors further demonstrated that this was the case when both the number of emotional dreams and emotions per dream were analyzed. Additionally, 8 (out of 10) discrete positive emotions and 5 (out of 10) discrete negative emotions were rated more frequently with SR than with ER. No gender differences were reported.

It should be noted that Kahan and LaBerge (1996) also compared ER and SR of emotions in dreams and dream reports. However, their aim was to compare dreaming and waking experiences with respect to certain cognitive and metacognitive processes, not emotions per se. Similarly to the two studies described above, the authors found a higher incidence of emotion when participants themselves rated their dream experiences than when judges rated the participants' dream reports (93.0% and 38.0%, respectively).

In sum, results from all of the studies converge in suggesting that with SR, as compared to ER, dreams are rated to be more emotional and relatively more positive. Also, the two methods differ more in the estimation of positive dreams and positive emotions than in the estimation of negative dreams and negative emotions. Results diverge with regard to the emotional valence of dreams: In the study by Schredl and Doll (1998) the majority of dreams were rated as negatively valenced with both SR and ER, whereas in the study by Sikka et al. (2014) the majority of dreams were rated as positively valenced with SR but to have a more balanced emotional tone with ER. Thus, in Sikka et al., as compared to Schredl and Doll, dreams were rated to be more positive with both methods of measurement.

The higher ratings of dream positivity with both SR and ER in Sikka et al. (2014) may result from methodological differences in the dream report collection environment and procedure. Specifically, whereas Schredl and Doll (1998) used written home dream reports collected upon spontaneous morning

awakenings, Sikka et al. collected verbal laboratory dream reports upon serial rapid eye movement sleep (REM) awakenings. Studies have shown that written dream reports collected upon home awakenings are more negatively valenced and contain more negative emotions than verbal laboratory dream reports collected upon serial REM awakenings, as measured with ER (Foulkes, 1979, Study 4) and SR (St-Onge et al., 2005). Additionally, these divergent results may reflect the different emotion rating scales used and differences in sample characteristics.

The Present Study: Aim and Hypotheses

The current study is a conceptual replication and extension of the laboratory study of Sikka et al. (2014). The aim is to compare the SR and ER of dream emotional experiences in a set of dreams collected in the home environment upon spontaneous morning awakenings (as in Schredl & Doll, 1998). To hold other methodological variables constant and ensure comparability with the previous study, the same scale as in Sikka et al. was used for measuring dream emotional experiences with both SR and ER. Thus, this study helps clarify the degree of convergence between the two methods for measuring dream emotional experiences. Additionally, it helps elucidate to what extent the negativity or positivity of dreams can be explained by who rates emotions (SR vs. ER) and to what extent by the dream report collection environment and procedure (morning awakenings at home vs. serial awakenings throughout the night in the laboratory).

Based on previous studies, it was hypothesized that (1) with SR, as compared with ER, dreams are estimated to be more emotional (i.e., a larger number of dreams are rated as emotional and to contain a larger number of emotions per dream); (2) the difference between SR and ER is larger for positive dreams and positive emotions in dreams than for negative dreams and negative emotions in dreams; and (3) with ER dreams are estimated to be more negative than positive and to contain more negative than positive emotions per dream, whereas with SR a more balanced ratio of positive and negative dreams, as well as positive and negative emotions per dream, is expected. Additionally, exploratory analyses regarding discrete emotions and gender differences were carried out without any specific prior predictions.

STUDY

METHOD

Participants and Procedure

Forty-seven nonpaid participants were recruited through advertisements sent to student mailing lists at the University of Skövde, Sweden, posted in Facebook and on a local online news site, and by contacting volunteers who had expressed an interest in participating in research studies. Prospective participants were told that the study concerned the content of dreams and the relation between dreams and well-being, but they remained unaware of the specific aims and research questions of the study. After being informed about the study process and signing the informed consent form, participants were first asked to fill in an online well-being questionnaire. It included measures of psychological and physical well-being, and sociodemographic variables, the results of which form part of another study and will be reported separately. Participants were then asked to log onto and fill in an online dream diary every morning upon awakening during the next three weeks (i.e., 21 days). To ensure there would be enough data from each participant, those with fewer than five dream reports over that period (one man and two women) were excluded from subsequent analyses. Thus, the final sample consisted of 44 participants (16 men, 28 women) with an average age of 26.9 years (SD = 5.1, range = 19–40). The mean age of men (M = 27.75, SD = 6.07) and women (M = 26.46, SD = 4.50) was not significantly different, $t(42)$ = 0.80, p = .427, d = 0.24. All participants were native Swedish speakers, were healthy, and reported not having been diagnosed with any psychiatric or neurological disorders. The study protocol and measures were approved by the regional ethical review board in Gothenburg, Sweden.

Home Dream Diary

Instructions for the dream diary were based on our previous study (Sikka et al., 2014). Participants were asked to systematically write down all the dreams they remembered during the next 21 days. Specifically, they were asked to keep paper and a pen next to their bed so that they could every morning upon awakening, but before getting up, quickly write down their dreams (so as to prevent forgetting as much as possible). Then, after getting up, participants were asked to log onto an online dream diary. First, they

were asked to indicate whether they remembered having any dreams that night by choosing one of the following answer options: "Yes, and I remember (at least some of) the contents," "Yes, I think I had a dream but I remember nothing about it," or "No, I think I had no dreams tonight." If they chose the latter two options, participants could end the session and submit the dream diary. If participants remembered having had a dream, they were instructed, "Write down the dream in as much detail as you can remember (what happened, where, when, who was present, what you felt and thought)." Such wording was used to ensure participants would report everything about their dream experience, including their emotional experiences, without directing their attention to only some aspects of the experience (i.e., emotions) and thereby exerting a possible influence on the content of the dream report. Participants were asked to report all details, even if they considered them to be unimportant and insignificant. They were also asked to not censor, change, or interpret the dream content or to make their dreams sound more reasonable and logical than the way they remembered them. In case they wanted to comment on some aspects of the dream experience, they were instructed to do so by adding the comments in brackets or at the end of the dream report in such a manner that they were clearly separable from the description of the actual dream experience.

After having reported the dream, participants were asked to rate the emotions they experienced in the dream by using the emotion rating scale. The latter was presented in the same online dream diary below the dream report field. In case participants remembered several dreams from the night, they were asked to write down the dream report and rate the dream emotions separately for each dream. The online dream diary was used to ensure participants would not retrospectively change their dream reports or emotion ratings, to provide better reports, and to keep track of whether participants logged onto the diary each day (in case they did not, e-mail reminders were sent).

Measures

SELF-RATINGS.

For purposes of comparability and consistency with the previous study (Sikka et al., 2014), the Swedish version of the modified Differential Emotions Scale (smDES; Fredrickson, 2013) was used to rate the emotions participants experienced in their dreams.

The smDES consists of 20 items or categories (each described by three adjectives), 10 for positive emotions and 10 for negative emotions. Participants rated each emotion item or category on a Likert scale from 0 ("I did not experience any of these feelings at all") to 4 ("I experienced one or more of these feelings extremely much").

To compare SR with ER, the frequency of occurrence, rather than the intensity, of emotions was considered. Therefore, the data gathered using the smDES were analyzed dichotomously, with a cutoff point between 0 (the emotion experienced "not at all") and 1 (the emotion experienced at least "a little bit"). Thus, the analyses were based on whether each emotion category was rated to occur or to not occur in a dream. Two subscales were formed separately for positive emotions (PE; $\alpha = .93$) and negative emotions (NE; $\alpha = .85$) by summing up the 10 PE items and 10 NE items, respectively.

EXTERNAL RATINGS.

All collected home dream reports were combined, randomized, and anonymized. Two judges content analyzed the dream reports, following the same criteria, procedure, and measures as in Sikka et al. (2014). First, the judges independently identified all instances of emotion that were explicitly expressed (E) by the dream self (e.g., "My sister and I got really scared"), that could be clearly inferred (I) from the behavior of the dream self (e.g., "My best friend had just come home from somewhere and we were laughing a lot about something I do not remember"), or both (EI) simultaneously (e.g., "I shouted because I was so angry").

Altogether, 581 emotions were identified in 552 dream reports. The judges agreed on 459 and disagreed on 122 instances of emotions. Thus, the identification interrater percent agreement was 79.0%. After discussion, the judges agreed on 570 instances of emotions (98.1%), and the 11 (1.9%) ambiguous cases were discarded. Of the 570 identified emotions, 457 (80.2%) were E, 67 (11.8%) I, and 46 (8.1%) EI. The interrater percent agreement for E was 84.7%, for I was 53.7%, and for EI was 73.9%. Because the interrater agreement for the 67 I emotions was low, they were excluded, and subsequent analyses were based on a total of 503 explicitly expressed (i.e., 457 E and 46 EI) emotions.

Next, the judges independently classified each of the identified emotions using the smDES. An additional category ("Other") was used for emotions the judges were not able to classify into any of the

20 emotion categories. In total, 42 instances of emotion (8.3%) were classified in this category: *surprised* (14), *confused* (9), *envious* (3), *relieved* (3), *bizarre* (2), *helpless* (2), *uncomfortable* (2), *sexually aroused* (1), *bored* (1), *cozy* (1), *dislike* (1), *missing* (1), *unstimulated* (1), and *fragmented* (1). For interrater reliability Cohen's κ (Landis & Koch, 1977) was computed using the classifications (i.e., scores of the 21 emotion categories) given by the two judges to each instance of emotion. There was almost perfect agreement (κ = .92) between the two judges' classifications.

The judges rated only the occurrence, not the intensity, of emotions. To enable comparisons with SR, the analyses were based on whether each emotion category was rated to occur or to not occur in a dream report (i.e., the same emotion category was considered only once even when it was scored several times in the same dream report). As with SR, two subscales were created for PE (10 items) and NE (10 items). The "Other" category was considered separately from these two subscales.

EMOTIONALITY OF DREAMS.

With SR, a dream was considered emotional if at least 1 of the 20 categories of the smDES was rated to occur at least once (i.e., received a score above 0) in a dream. With ER, a dream was considered emotional if at least 1 of the 21 categories (20 categories of smDES plus the category "Other") was detected at least once in a dream report. The category "Other" was included to ensure that all the dream reports in which emotions were identified would be rated as emotional.

EMOTIONAL VALENCE OF DREAMS.

A dream was considered as having a balanced emotional valence when the frequencies of positive and negative emotion categories were equal; otherwise it was regarded as either positive (more positive than negative emotion categories) or negative (more negative than positive emotion categories). Additionally, a dream report containing only the emotions that with ER were classified into the "Other" category were considered as having an undetermined valence (because of the ambiguous nature of these emotion items).

POSITIVE AND NEGATIVE EMOTIONS PER DREAM.

The number of positive emotions and negative emotions per dream were obtained by summing up the occurrence of the 10 positive emotion and 10 negative emotion categories, respectively. The overall number of emotions was obtained by summing up the PE

and NE subscales. Thus, the maximum number of both PE and NE per dream was 10 and of all different emotions per dream 20.

Length of Dream Reports

The length or word count of dream reports was calculated according to Antrobus (1983), including all dream-related words but excluding repetition, fillers, corrections, and waking comments.

Statistical Analyses

All analyses reported below were conducted using the subject-level aggregation. Dream-level analyses were used only for calculating the percentage of nonemotional, emotional, and emotionally valenced dreams and for analyzing the co-occurrence of different discrete emotions in the same dream.

IBM SPSS statistics software (version 20) was used to conduct statistical analyses. The Shapiro–Wilk test (Shapiro & Wilk, 1965) was used to test the normality assumption. Because the majority of the variables were not normally distributed, most comparisons were conducted using nonparametric tests (Mann–Whitney U test or Wilcoxon signed-rank test). Parametric tests (paired-samples t test) were used when both variables in the comparison were normally distributed. Effect sizes were calculated using Pearson's correlation (r) and with Cohen's d, respectively. All statistical tests were two-tailed, and for nonparametric tests exact tests were conducted. Correlation analyses were conducted using the Spearman's rank correlation coefficient (r_s).

RESULTS

A total of 552 (M = 12.55, Mdn = 11, SD = 5.72) dream reports were provided over the 3-week period.

Emotionality of Dreams

With SR, 538 (97.5%) dreams, whereas with ER 264 (47.8%) dream reports, were rated as emotional (Table 1). With SR every participant had at least five emotional dreams, with an average of 12.23 emotional dreams per participant. With ER four participants did not have any emotional dream reports, with an average of 6.00 emotional dream reports per participant. Thus, with SR a significantly larger number of dreams were rated to be emotional than with ER, with a large effect size of r = .61 (see Table 2 for descriptive statistics and significance tests).

TABLE 1. Percentage of Nonemotional and Emotionally Valenced Dreams as Measured With Self- and External Ratings in the Home (Current Study) and Laboratory Environment (Sikka et al., 2014)

| | Home morning dreams (N = 552) | | | | Laboratory REM dreams[a] (N = 115) | | | |
| | Self-ratings | | External ratings | | Self-ratings | | External ratings | |
	% All	% Emotional	% All	% Emotional	% All	% Emotional	% All	% Emotional
Nonemotional dreams	2.5		52.2		0.0		71.3	
Emotional dreams	97.5		47.8		100.0		28.7	
Positive dreams	55.8	57.2	12.5	26.1	79.1	79.1	9.6	33.4
Negative dreams	35.3	36.2	28.1	58.7	12.2	12.2	11.3	39.4
Balanced dreams	6.3	6.5	5.1	10.6	8.7	8.7	1.7	5.9
Undetermined[b]	N/A	N/A	2.2	4.5	0.0	0.0	6.1	21.3

[a]Results from Sikka et al. (2014).
[b]Dream reports containing only emotions classified by judges in the "Other" category.

TABLE 2. Mean and Median Number of Emotionally Valenced Dreams and Emotions per Dream as Measured with Self- and External Ratings

| | Self-ratings | | | External ratings | | | Wilcoxon signed rank test (2-tailed) | | Effect size |
	M	SD	Mdn	M	SD	Mdn	Z	p	r
Emotional dreams	12.23	5.60	11.00	6.00	4.29	5.00	−5.72	<.001	.61
Positive dreams	7.00	4.63	6.00	1.57	1.89	1.00	−5.54	<.001	.59
Negative dreams	4.43	3.45	4.00	3.52	3.24	3.00	−2.36	=.017	.25
Balanced dreams	0.80	1.15	0.00	0.64	0.75	0.00	−0.74	=.476	.08
Undetermined[a]	N/A	N/A	N/A	0.27	0.62	0.00	N/A	N/A	N/A
Emotions per dream	8.45	3.08	8.58	0.76[b]	0.61[b]	0.73[b]	−5.78	<.001	.62
Positive emotions	4.89	2.09	4.76	0.27	0.27	0.20	−5.78	<.001	.62
Negative emotions	3.57	1.79	3.75	0.49	0.42	0.38	−5.78	<.001	.62

[a]Dream reports containing only emotions classified by judges in the "Other" category (12 dreams of 8 participants).
[b]Excludes emotions classified in the "Other" category.

Emotional Valence of Dreams

Table 1 presents the percentages of dreams with different emotional valence for both SR and ER, and Table 2 presents the mean and median number of emotionally valenced dreams per participant together with corresponding significance tests.

Whereas with SR participants reported more positive than negative dreams, Wilcoxon $Z = -2.91$, $p = .003$, $r = .31$, with ER judges rated a larger number of dream reports to be negative than positive, Wilcoxon $Z = -3.59$, $p < .001$, $r = .38$. When the two measures were compared directly, significantly more dreams were rated positive with SR than with ER, with a large effect size of $r = .59$. Although significantly more dreams were also rated negative with SR, as compared to ER, the number of negative dreams was rather similar with both measures and the effect size small ($r = .25$) (Figure 1).
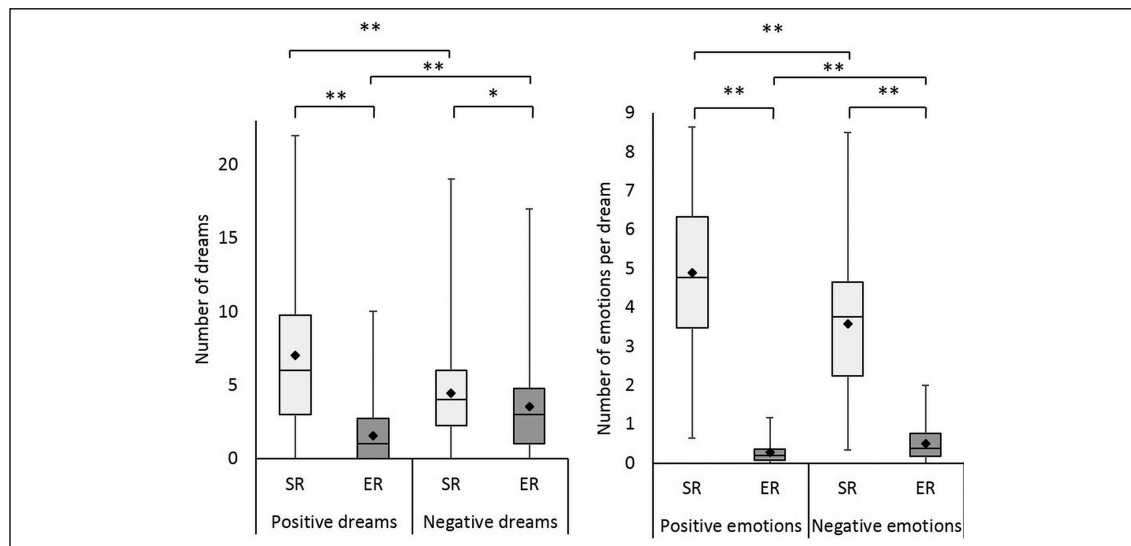
FIGURE 1. Differences between self-ratings (SR) and external ratings (ER) in the number of positive and negative dreams and positive and negative emotions per dream. ♦ = mean; horizontal line = median; box = 25th and 75th percentiles; whiskers = minimum and maximum values. Lines over bars indicate significant differences. **$p < .01$; *$p < .05$

Positive and Negative Emotions per Dream

With SR participants reported significantly more PE than NE per dream, paired-samples $t(43) = 3.65$, $p = .001$, $d = 0.55$. In contrast, with ER participants' dream reports were rated to contain more NE than PE, Wilcoxon $Z = –3.64$, $p < .001$, $r = .39$. When the two measures were compared, SR reflected significantly more emotions per dream in general and more PE and NE per dream than ER, with large effect sizes ($r = .62$) (see Figure 1 and Table 2 for descriptive statistics and significance tests). However, the difference between SR and ER was significantly larger for PE ($M = 4.62$, $SD = 2.09$, $Mdn = 4.45$) than for NE ($M = 3.08$, $SD = 1.68$, $Mdn = 3.19$), paired-samples $t(43) = 4.50$, $p < .001$, $d = 0.68$. The pattern of results was the same when only emotional dreams (with both measures) were included in the analyses.

There was a positive correlation between the number of different emotion categories per dream and the length of the dream report with both measures. Specifically, the mean word count of the dream report ($M = 122.95$, $Mdn = 103.66$, $SD = 69.99$) was moderately to strongly correlated with the total number of emotions per dream (SR: $r_s = .47$, $p = .001$; ER: $r_s = .35$, $p = .021$), the number of PE per dream (SR: $r_s = .33$, $p = .028$; ER: $r_s = .37$, $p = .014$), and the number of NE per dream (SR: $r_s = .41$, $p = .006$; ER: $r_s = .34$,

$p = .023$). Hence, the longer the dream report, the more emotions were rated with both measures.

With SR the PE and NE scales were not correlated with each other ($r = .25$, $p = .106$), whereas with ER there was a positive correlation between PE and NE ($r_s = .49$, $p = .001$). Moreover, whereas the NE subscales of the two measures were positively correlated ($r_s = .48$, $p = .001$), the PE subscales were not significantly correlated ($r_s = .22$, $p = .152$). Hence, the ratings of NE with SR and ER correspond better to each other than the ratings of PE.

Exploratory Analyses

DISCRETE EMOTIONS PER DREAM.
The comparison of SR and ER with regard to the occurrence of all the different discrete emotions per dream were based on only the dreams that were rated emotional with both SR and ER ($n = 264$). The mean and median occurrence of each of the discrete emotion categories per emotional dream (in the order of the most to the least frequent) with both SR and ER, together with statistical test results, are presented in Table 3. As can be seen, all discrete emotions were rated significantly more frequently with SR than with ER, with large effect sizes ($r = .45$-.60).

With both measures the same discrete emotions were among the most frequently rated categories: *joy-*

TABLE 3. Mean and Median Number of Discrete Emotions per Emotional Dream as Measured with Self- and External Ratings

	Self-ratings			External ratings			Wilcoxon signed rank test (2-tailed)		Effect size
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*Z*	*p*	*r*
Positive emotions									
Joyful/glad/happy	0.60	0.26	0.60	0.13	0.15	0.10	−5.40	<.001	.60
Interested/alert/curious	0.59	0.30	0.60	0.09	0.13	0.00	−5.09	<.001	.57
Amused/fun-loving/giggly	0.54	0.29	0.57	0.03	0.07	0.00	−5.24	<.001	.59
Hopeful/optimistic/encouraged	0.54	0.31	0.54	0.01	0.04	0.00	−5.24	<.001	.59
Love/closeness/trust	0.51	0.30	0.50	0.03	0.07	0.00	−5.16	<.001	.59
Serene/content/peaceful	0.48	0.25	0.50	0.07	0.11	0.00	−5.25	<.001	.59
Inspired/uplifted/elevated	0.48	0.29	0.50	0.01	0.05	0.00	−5.09	<.001	.56
Awe/wonder/amazement	0.47	0.32	0.50	0.04	0.10	0.00	−5.04	<.001	.56
Proud/confident/self-assured	0.47	0.27	0.50	0.04	0.11	0.00	−5.09	<.001	.58
Grateful/appreciative/thankful	0.40	0.33	0.40	0.06	0.12	0.00	−4.20	<.001	.46
Negative emotions									
Stressed/nervous/overwhelmed	0.64	0.30	0.72	0.19	0.18	0.19	−5.24	<.001	.54
Angry/irritated/annoyed	0.57	0.35	0.60	0.24	0.24	0.20	−4.40	<.001	.49
Scared/fearful/afraid	0.54	0.27	0.50	0.18	0.18	0.18	−4.87	<.001	.54
Embarrassed/self-conscious/blushing	0.47	0.31	0.48	0.05	0.10	0.00	−5.01	<.001	.56
Sad/downhearted/unhappy	0.46	0.31	0.45	0.15	0.21	0.10	−4.74	<.001	.50
Ashamed/humiliated/disgraced	0.44	0.27	0.48	0.01	0.04	0.00	−5.16	<.001	.58
Disgust/distaste/revulsion	0.43	0.30	0.48	0.03	0.07	0.00	−4.87	<.001	.54
Hate/distrust/suspicion	0.39	0.29	0.35	0.04	0.09	0.00	−4.85	<.001	.54
Guilty/repentant/blameworthy	0.37	0.25	0.33	0.05	0.11	0.00	−5.00	<.001	.55
Contemptuous/scornful/disdainful	0.18	0.23	0.15	0.01	0.03	0.00	−4.16	<.001	.45

ful/glad/happy and *interested/alert/curious* among positive emotion categories and *stressed/nervous/overwhelmed, angry/irritated/annoyed,* and *scared/fearful/afraid* among negative emotion categories. Similarly, *contemptuous/scornful/disdainful* was the least frequently rated emotion category with both measures.

Analyses including all dreams (N = 552) showed that with SR all 10 PE categories were positively correlated with each other, as was the case for NE categories. However, intercorrelations (rs) between the PE categories were stronger (range = .28–.62, M = .46, SD = .08) than those between the NE categories

(range = .12–.59, M = .33, SD = .11), paired-samples $t(44)$ = 5.81, p < .001, d = 0.98. This means that participants were more likely to rate the co-occurrence of several different PE than NE categories in the same dream. With ER, only a few PE and NE categories were intercorrelated, and the correlations did not exceed rs = .19.

GENDER DIFFERENCES.
Comparing men and women, the only significant differences were found when dream reports were externally rated. Specifically, women expressed more NE in their dream reports and, as a result, were rated to

TABLE 4. Gender Differences Regarding the Mean and Median Number of Emotionally Valenced Dreams and Emotions per Dream as Measured With Self- and External Ratings

	Men (*n* = 16)			Women (*n* = 28)			Mann–Whitney or independent samples *t* test (2-tailed)			Effect size
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*U* or *t*[a]	*Z*	*p*	*r* or *d*[a]
Self-ratings										
Emotional dreams	11.81	5.31	11.50	12.46	5.85	10.50	213.50	−0.26	.804	.04
Positive dreams	7.63	5.50	6.00	6.64	4.11	6.50	213.00	−0.27	.795	.04
Negative dreams	3.38	2.31	3.00	5.04	3.86	4.00	164.00	−1.48	.143	.22
Balanced dreams	0.81	1.33	0.00	0.79	1.07	0.00	212.50	−0.33	.761	.05
Emotions per dream	8.21	2.96	8.74	8.59	3.18	8.50	−0.40[a]	N/A	.693	0.12[a]
Positive emotions	4.99	2.13	5.04	4.82	2.11	4.54	0.26[a]	N/A	.799	0.08[a]
Negative emotions	3.21	1.78	2.82	3.77	1.80	3.88	−0.99[a]	N/A	.327	0.31[a]
External ratings										
Emotional dreams	4.56	3.81	4.50	6.82	4.40	6.00	156.50	−1.66	.099	.25
Positive dreams	1.88	2.42	1.00	1.39	1.52	1.00	201.50	−0.57	.577	.09
Negative dreams	2.13	1.54	2.00	4.32	3.68	4.00	136.00	−2.17	.029	.33
Balanced dreams	0.44	0.63	0.00	0.75	0.80	1.00	177.50	−1.25	.240	.19
Undetermined[b]	0.13	0.34	0.00	0.36	0.73	0.00	200.00	−0.87	.438	.13
Emotions per dream[c]	0.64	0.75	0.45	0.83	0.52	0.79	149.50	−1.82	.069	.27
Positive emotions	0.27	0.31	0.18	0.27	0.25	0.21	211.50	−0.31	.766	.05
Negative emotions	0.37	0.49	0.22	0.56	0.37	0.59	135.50	−2.16	.030	.33

[a]*Independent samples* t *test*, N = 44, df = 42.
[b]*Dream reports containing only emotions classified by judges in the "Other" category (12 dreams of 8 participants).*
[c]*Excludes emotions classified in the "Other" category.*

have more negative dream reports than men (Table 4). There were no differences between men and women in any of the variables when they rated their dream emotions themselves. Despite women having expressed more NE in their dream reports and having more negative dream reports (as measured with ER), both men and women displayed the same differences between SR and ER described above.

Interestingly, however, when the differences between positive and negative dream reports and PE and NE per dream report (as measured with ER) were analyzed separately for the two genders, a different pattern of results occurred. Women were rated to have more negative than positive dream reports, Z = −3.54, p < .001, r = .47, and more NE than PE per

dream report, Z = −3.50, p < .001, r = .47. For men, on the other hand, the number of negative and positive dreams reports, Z = −1.32, p = .203, r = .23, and NE and PE per dream report, Z = −1.02, p = .336, r = .18, was balanced and not significantly different.

The higher number of NE in women's dream reports could not be explained by the length of the report because it did not significantly differ between those of men (n = 16; M = 112.85, SD = 56.82, Mdn = 102.53) and women (n = 28; M = 128.73, SD = 76.89, Mdn = 106.15), Mann–Whitney U = 207.00, Z = −0.42, p = .690, r = .06.

Importantly, whereas for women the NE scales of SR and ER were strongly positively correlated (r_s = .70, p = .000), for men the correlation was not signifi-

cant (r_s = .12, p = .653). Thus, the way women rated their negative dream emotions corresponded to how they expressed NE in their dream reports, but this was not the case for men.

DISCUSSION

With Self-Ratings Dreams Appear to Be More Emotional Than With External Ratings

There were large differences in dream emotionality obtained with the two measures. SR, as compared to ER, resulted in twice as many estimates of emotional dreams and in more than 10 times more estimates of different emotions per dream. Moreover, all the discrete emotions were rated to occur more frequently with SR than with ER. Hence, with SR dreams are estimated to be more emotional than with ER. These findings confirm Hypothesis 1 and are well in line with previous studies (Kahan & LaBerge, 1996; Schredl & Doll, 1998; Sikka et al., 2014).

With Self-Ratings Dreams Appear to Be More Positive Than With External Ratings

With SR more than half of the dreams, whereas with ER only about one tenth of the dreams, were found to be positively valenced. Thus, SR resulted in more than four times more estimates of positive dreams than ER. In contrast, with both measures one third of the dreams were found to be negatively valenced. Although the difference between the measures with regard to the number of negative dreams was significant at the .05 level, it was small and occurred only toward the end of the study period (see Supplementary Results). Regarding the number of emotions per dream, SR resulted in significantly more estimates of both PE and NE than ER. However, the difference between SR and ER was more marked with regard to PE than NE. The results confirm Hypothesis 2 and suggest that the differences between the two measures lie mostly in the number of positive dreams and positive emotions in a dream. This was further corroborated by correlation analyses, which demonstrated that the ratings of negative emotions of the two measures correspond more to each other than the ratings of positive emotions. Similar conclusions were reached in both Schredl and Doll (1998) and Sikka et al. (2014).

With Self-Ratings Dreams Appear to Be Mostly Positive, With External Ratings Mostly Negative

Interestingly, whereas SR reflected more positive than negative dreams and more PE than NE per dream, with ER the results were exactly the opposite: more negative than positive dreams and more NE than PE per dream. Therefore, SR can lead to the conclusion that dreams are mostly positive, whereas ER can lead to the conclusion that dreams are mostly negative. However, the negativity of dreams with ER characterized the dream reports of the women only, because they expressed more NE in their dream reports. The men, on the other hand, expressed PE and NE in their dream reports rather equally. Thus, Hypothesis 3 was partly confirmed.

The differences in the negative emotional content of dream reports of men and women were surprising because previous studies have typically not reported such differences, at least when content analysis has been used (e.g., Hall & Van de Castle, 1966; Merritt et al., 1994; Schredl, 2002; Schredl & Doll, 1998). However, in a recent study among a large sample of Canadian university students, women also expressed a higher percentage of NE in their dream reports than men (Dale, Lortie-Lussier, Wong, & De Koninck, 2016). The strong positive correlation between self-rated NE and NE expressed in corresponding dream reports in women indicates that their NE ratings with the two methods show better convergent validity and suggests that men might have underreported NE in their dream reports. The reason for such underreporting is unclear, but it might reflect differences in how men and women express NE in language. It is known that the use of emotional language is related to personality traits, such that people high in neuroticism use more NE words (see Ireland & Mehl, 2014, for a review). However, in a study by Holtgraves (2010) on language use in everyday text messages, this relationship was found for women but not for men. Additionally, research suggests that men tend to describe NE, as compared to PE, in more nonliteral language, whereas no such pattern has been found for women (Link & Kreuz, 2005). This seems to suggest gender differences in socialization and emotional display rules, as a result of which men might regulate their emotional language to a greater extent than women and be less willing to disclose NE, or at least

do so using more nonliteral language (Brody & Hall, 2016; Ireland & Mehl, 2014; Link & Kreuz, 2005).

Nevertheless, in the present study the ratio of positive to negative dreams with SR was indeed smaller (or more balanced) than in the laboratory study of Sikka et al. (2014) (see Table 1 for the comparison of the results from the two studies). This can be explained by the smaller percentage of positive dreams and the higher percentage of negative dreams in the current study. With regard to ER, the ratio of negative to positive dreams was larger here than in Sikka et al. (2014) because of the higher percentage of negative dreams. These findings provide further evidence that dream reports collected at home upon morning awakenings are indeed less positive and more negative than dream reports collected in the laboratory upon serial REM awakenings (see also Foulkes, 1979, Study 4; St-Onge et al., 2005).

Theoretical and Empirical Implications of the Present Study

This study demonstrates that the use of different methods—ER of narrative dream reports versus SR using emotion rating scales—can lead to strikingly different results and contradictory conclusions about emotional experiences in dreams. Similar discrepancies have been observed in the study of waking state emotional experiences. Specifically, Kahan and LaBerge (1996) found that judges rated a much lower incidence of emotions in waking narrative reports (35%) as compared to when participants themselves rated their waking emotions using a questionnaire (83%). To our knowledge, surprisingly little research has been published on the same issue outside dream research. Among the few studies that have concurrently used the two measures in the study of waking state emotional experiences, findings show only partial convergence between self-rated emotions and emotions expressed in narrative reports analyzed by either external raters (e.g., Suslow, Battacchi, & Renna, 1996) or computerized content analysis programs (e.g., Kahn, Tobin, Massey, & Anderson, 2007; Tov, Ng, Lin, & Qiu, 2013). Thus, the lack of correspondence between the two methods is an issue relevant not only for dream research but for emotion research in general. The question arises as to what might explain such large differences in results obtained with the two methods. Below the various conceptual and

methodological issues are discussed, and recommendations as to how to address them are provided.

First, the two methods represent two different types of self-reports that result in two different sets of data: ratings of emotional experiences and emotional language use in narrative reports. Thus, the differences are not only in *who* rates emotional experiences but *what* is being rated (Sikka et al., 2014).

Second, the temporal unit of analysis differs between the two methods. Research on the memory–experience gap has demonstrated that the average of experienced emotions and the retrospective evaluation of the same experience as a whole differ, with the latter reflecting a larger number of positive and negative emotional experiences (Miron-Shatz, Stone, & Kahneman, 2009; but see also Miron-Shatz, 2009, for contrasting evidence). This results from memory bias and aggregation processes (i.e., respondents have to combine multiple experiences into one overall report) (Larsen & Fredrickson, 1999). Although both SR of dream experiences and narrative dream reports are necessarily retrospective, they may be subject to different types of memory biases. However, differently from narrative reports, SR typically ask respondents to rate the dream as a whole (i.e., aggregate experiences across the whole dream) rather than dream events or episodes (although in a few dream studies participants were asked to rate emotions in their own dream reports on a line-by-line basis; e.g., Fosse et al., 2001; Merritt et al., 1994; or separately for each scene; Nielsen et al., 1991).

Third, the two methods may actually measure different feeling states: emotions and moods. Although they are closely related, the difference between these states is that whereas emotions occur as a response to and focus on specific external or internal events, objects, or people, are intense, and are short-lasting (seconds or minutes), moods lack a specific focus, are more diffuse, have a lower felt intensity, and last longer (hours or days) (Juslin & Västfjäll, 2008; Scherer, 2005). It may well be that the storylike structure of narrative reports provides a context and may make it more likely that emotions, rather than moods, are reported (i.e., how one felt as a result of something that happened). With SR, being less constrained by

the reported content of the dream, participants may be more likely to also take into account the general mood of the dream. Research has shown that people tend to experience mildly positive mood most of the time, even when no particular emotional events are taking place, a phenomenon known as the positivity offset (Ashare, Norris, Wileyto, Cacioppo, & Strasser, 2013; Cacioppo, Gardner, & Berntson, 1999). It has been suggested that people might report positive moods not because of their presence but because of the absence of negative emotions or moods (Diener, Kanazawa, Suh, & Oishi, 2015). Thus, it may well be that because of the positivity offset, SR of emotional experiences are biased in the positive direction (see Table S3 in the Supplementary Results for examples of dream reports).

Fourth, it has been argued that narrative reports may be influenced by several individual differences (e.g., participants' language skills, motivation, awareness of feelings; Kahan, 2012) and by selective reporting of the content (i.e., participants are more likely to report what, when, who, and where rather than feelings and thoughts) (Kahan, 1994; Kahan & LaBerge, 1996, 2011). Thus, emotions may be less likely to be included in the narrative (dream) report, and as a result these may be incomplete representations of (dream) experiences (Kahan & Horton, 2012). Moreover, it has been suggested that the reason why ER reflect less positive emotions in dream reports may be our linguistic tendencies to express negative experiences in specific emotions but positive experiences either by general moods (Strauch & Meier, 1996) or implicitly by describing situations rather than how they made us feel (Sikka et al., 2014) (see Table S3 in the Supplementary Results for examples of dream reports). Thus, ER of dream content may be less able to capture such positive emotional experiences.

Although some have argued that SR more validly represent dream emotional experiences (Kahan & Claudatos, 2016; Schredl, 2010b; Strauch & Meier, 1996), others have pointed out that such SR may also be biased, for example, by the demand characteristics of the rating task (i.e., participants may be primed by answer options) or waking state cognition (i.e., certain emotions may be assumed to have been present in certain dream situations even if they were not) (Domhoff, 2005; Zadra & Domhoff, 2017). Moreover,

PE may be more diffuse and blended than NE (Mikels et al., 2005), as a result of which participants may rate several discrete emotions to characterize one positive emotional experience. The finding that intercorrelations between discrete PE categories were stronger than those for NE categories suggests that PE are indeed less differentiated (Ellsworth & Smith, 1988).

In sum, the two rating methods may differ in their estimation of (dream) emotionality because of the strengths and weaknesses inherent in both methods. SR by participants using emotion rating scales may reflect a broader range of feeling states, that is, moods and implicit states that are otherwise difficult to directly express. At the same time such ratings may be biased by the phenomenon of positivity offset, aggregation processes, demand characteristics of the rating task, and the diffuse nature of positively valenced states, leading to the overestimation of certain, especially positive, emotional experiences. ER, on the other hand, typically measure only specific emotions that are directly expressed in the narrative report, which ensures reliability and replicability of the results. However, because of selective reporting and individual or linguistic differences in expressing emotions, ER of narrative reports may lead to the underestimation of certain, especially positive, emotional experiences.

RECOMMENDATIONS FOR FUTURE STUDIES.
As this discussion indicates, there are good arguments for and against the validity of either method. At the moment, there is no agreement as to which method should be preferred. In dream research, content analysis of narrative reports has been traditionally used, but recently a case for using SR has been made (e.g., Kahan & Claudatos, 2016). Outside dream research, SR scales have been the traditional way of measuring subjective emotional experience; however, recently researchers have become more critical of such an approach (e.g., Heavey et al., 2012; Hurlburt & Heavey, 2015; Scherer, 2005). Therefore, it is important to conduct more convergent validity studies involving both dream and nondream experiences and narratives and to minimize possible biases inherent in both methods and to collect data under ideal reporting conditions (see also Windt, 2013, 2015).

First, to distinguish the effect of the rater (who rates) from the effect of the rating task (what is rated),

a direct comparison of SR and ER of (emotional experiences in) narrative reports is needed.

Second, to minimize the memory–experience gap and aggregation of experiences, it is important to use dense and frequent measurement of momentary experiences (i.e., experience sampling) and to measure specific (dream or waking) episodes or scenes rather than aggregate experiences over a longer period of time.

Third, it is important to acknowledge the conceptual difference between emotions and moods. Although this distinction is theoretically recognized, it is not represented in the methods and measurements used in empirical (dream or emotion) research. Therefore, "separating positive emotions from moods is an important task for future research" (Diener et al., 2015, p. 249). Following conceptual distinctions could be a fruitful way to approach this problem and thus to measure not only the intensity (i.e., whether the emotional experience had a high, moderate, or low intensity) but also the duration (i.e., how long the particular emotional experience lasted) and context (i.e., whether the emotional experience was a reaction to, and focused on, some event that occurred) of the emotional experience.

Fourth, training of the participants in reporting and rating their emotional experiences is increasingly recognized as an important aspect in both dream (Kahan, 2012; Kahan & Claudatos, 2016; Windt, 2015) and emotion research (Heavey et al., 2012; Hurlburt & Heavey, 2015; Nielsen & Kaszniak, 2007). If we assume that self-reports provide valid accounts of subjective emotional experience, then we have to ensure their accuracy and completeness. Specifically, training is needed to develop awareness of one's feeling states and their difference from each other, language and reporting skills, to counteract demand characteristics, and the influence of waking state emotional and cognitive processes on the reported or rated experiences. However, it is important to be aware of the possible influence training in itself might have on the experience and reporting of phenomenal experiences (Windt, 2015).

Finally, it may well be that both methods provide equally valid results, and the contrasting findings can be explained by different types of data being obtained and different phenomena being measured (emotional experiences vs. emotional language; feeling states in general vs. emotions). Thus, both methods provide important information, and the choice of the method should be explicitly articulated by researchers based on the phenomenon of interest. If self- and external ratings measure or capture at least partially different phenomena, they should be considered complementary and used concurrently.

Possible Limitations of the Study

Because participants rated their emotional experiences in dreams after having written the narrative dream report, the results may have been affected by the order effect. On one hand, the report or reporting process may have interfered with dream memory and led to inaccurate ratings of the dream experience (Parke & Horton, 2009). On the other hand, participants may have inferred that because emotion ratings always followed narrative dream reports, they did not need to report everything, especially their emotional experiences, in too much detail, thus resulting in incomplete narrative reports. In future studies, counterbalancing the order of narrative reports and emotion ratings should help clarify this issue.

Additionally, to counteract the decay of the memory of the dream experience, to obtain better reports, and to ensure compliance with the study procedure, participants were instructed to think back to the dream and take notes of it on a piece of paper before getting out of bed, logging onto the online dream diary, and typing in their report. It may be argued that such "double reporting" may have altered the report because participants may have inadvertently modified the content of the dream in the process of transferring or typing it to the online dream diary. Given that emotion ratings followed the online dream report, the ratings may have also been altered by the temporal delay between the experience and the rating task. However, the correspondence of the results of the present study with those from Sikka et al. (2014), where verbal dream reports were taped and emotion ratings conducted directly upon laboratory awakenings, speak against such influences.

Furthermore, it may be argued that because the time line of the study was long, differences in the results obtained with the two methods may have been affected by a motivational or attentional bias (Zadra & Robert, 2012). Specifically, participants' motivation in writing time-consuming narrative dream reports

may have decreased over time. In contrast, paying attention to emotional experiences (using the emotion rating scale) may have specifically affected the reporting or rating of such experiences. However, post hoc analyses regarding possible differences in the ratings of emotions over the study period (see Supplementary Results) showed that the length of the study, and hence the motivational and attentional factors, cannot explain the differences between SR and ER.

Conclusion

This study demonstrates that the use of different methods for measuring emotional experiences in dreams—ER of narrative dream reports and SR using emotion rating scales—can lead to strikingly different results. It provides further evidence that when participants themselves rate their dream emotional experiences using emotion rating scales, dreams appear to be more emotional and more positive than when narrative dream reports are content analyzed by external raters. Importantly, results obtained with the two methods can lead to conflicting conclusions: With SR dreams can be claimed to be mostly positively valenced, and with ER dreams can be claimed to be mostly negatively valenced. However, these conclusions depend not only on the measurement method but also on *whose* dream reports are evaluated. Whereas the externally rated dream reports of women were mostly negatively valenced, those of men were equally likely to be positively and negatively valenced. Whereas SR and ER converge, at least partially, in the measurement of negative emotional experiences, they diverge greatly in the measurement of positive emotional experiences. This suggests that negative emotional experiences can be measured in a more valid and reliable manner than positive emotional experiences. Thus, at present it is not possible to draw overall conclusions about dream emotionality (e.g., the negativity or positivity of dreams), but the conflicting results highlight a need for a more conceptually informed (e.g., how to better distinguish between emotions and moods) and methodologically sound (e.g., how to best capture moments when emotions are experienced) research.

NOTES

Diana Feilhauer is now at the Department of Philosophy, Lund University, Sweden.

Address correspondence about this article to Pilleriin Sikka, Centre for Cognitive Neuroscience, Turku Brain and Mind Center, Department of Psychology and Speech–Language Pathology, University of Turku, Finland. (e-mail: pilsik@utu.fi).

REFERENCES

Antrobus, J. (1983). REM and NREM sleep reports: Comparison of word frequencies by cognitive classes. *Psychophysiology, 20,* 562–568. http://dx.doi.org/10.1111/j.1469-8986.1983.tb03015.x

Ashare, R. L., Norris, C. J., Wileyto, E. P., Cacioppo, J. T., & Strasser, A. (2013). Individual differences in positivity offset and negativity bias: Gender-specific associations with two serotonin receptor genes. *Personality and Individual Differences, 55,* 469–473. doi:10.1016/j.paid.2013.04.009

Barrett, L. F. (1996). Hedonic tone, perceived arousal, and item desirability: Three components of self-reported mood. *Cognition & Emotion, 10,* 47–68. doi:10.1080/026999396380385

Blagrove, M., Framer, L., & Williams, E. (2004). The relationship of nightmare frequency and nightmare distress to well-being. *Journal of Sleep Research, 13,* 129–136. http://dx.doi.org/10.1111/j.1365-2869.2004.00394.x

Blick, K. A., & Howe, J. B. (1984). A comparison of the emotional content of dreams recalled by young and elderly women. *Journal of Psychology, 116,* 143–146. doi:10.1080/00223980.1984.9923629

Brody, L. R., & Hall, J. A. (2016). Gender and emotion in context. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones, *Handbook of emotions* (4th ed., pp. 395–408). New York, NY: Guilford.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76,* 839–855. http://dx.doi.org/10.1037/0022-3514.76.5.839

Cartwright, R. D. (1996). Dreams and adaptation to divorce. In D. Barrett (Ed.), *Trauma and dreams* (pp. 173–185). Cambridge, MA: Harvard University Press.

Dale, A., Lortie-Lussier, M., Wong, C., & De Koninck, J. (2016). Dreams of Canadian students: Norms, gender differences, and comparison with American norms. *Journal of Cross-Cultural Psychology, 47,* 941–955. doi:10.1177/0022022116655788

Diener, E., Kanazawa, S., Suh, E. M., & Oishi, S. (2015). Why people are in a generally good mood. *Personality and Social Psychology Review, 19,* 235–256. doi:10.1177/1088868314544467

Domhoff, G. W. (2003). *The scientific study of dreams: Neural networks, cognitive development, and content analysis.* Washington, DC: American Psychological Association.

Domhoff, G. W. (2005). The content of dreams: Methodologic and theoretical implications. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (4th ed., pp. 522–534). Philadelphia, PA: Elsevier.

Ellsworth, P. C., & Smith, C. A. (1988). Shades of joy: Patterns of appraisal differentiating pleasant emotions. *Cognition & Emotion, 2,* 301–331. http://dx.doi.org/10.1080/02699938808412702

Fosse, R., Stickgold, R., & Hobson, J. A. (2001). The mind in REM sleep: Reports of emotional experience. *Sleep, 24,* 1–9. doi:10.1111/1467-9280.00306

Foulkes, D. (1979). Home and laboratory dreams: Four empirical studies and a conceptual reevaluation. *Sleep, 2,* 233–251.

Fredrickson, B. L. (2013). Positive emotions broaden and build. In G. Devine & E. A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 1–53). San Diego, CA: Academic Press.

Hall, C. S. (1951). What people dream about. *Scientific American, 184,* 60–63.

Hall, C. S., & Van de Castle, R. L. (1966). *The content analysis of dreams.* New York, NY: Appleton-Century-Crofts.

Hartmann, E. (1995). Making connections in a safe place: Is dreaming psychotherapy? *Dreaming, 5,* 213–228. http://dx.doi.org/10.1037/h0094437

Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition, 17,* 798–810. doi:10.1016/j.concog.2007.12.006

Heavey, C. L., Hurlburt, R. T., & Lefforge, N. L. (2012). Toward a phenomenology of feelings. *Emotion, 12,* 763–777. doi:10.1037/a0026905

Holtgraves, T. (2010). Text messaging, personality, and the social context. *Journal of Research in Personality, 45*(1), 92–99. doi:10.1016/j.jrp.2010.11.015

Hsu, S. S., & Yu, C. K.-C. (2016). Content analysis of Chinese dreams: Pleasure or pain? *Dreaming, 26*(3), 208–220. http://dx.doi.org/10.1037/drm0000032

Hurlburt, R. T., & Heavey, C. L. (2015). Investigating pristine inner experience: Implications for experience sampling and questionnaires. *Consciousness and Cognition, 31,* 148–159. http://dx.doi.org/10.1016/j.concog.2014.11.002

Ireland, M. E., & Mehl, M. R. (2014). Natural language use as a marker of personality. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 201–218). New York, NY: Oxford University Press.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31,* 559–621. doi:10.1017/s0140525x08005293

Kahan, T. L. (1994). Measuring dream self-reflectiveness: A comparison of two approaches. *Dreaming, 4*(3), 177–193. doi:10.1037/h0094411

Kahan, T. L. (2012). Cognitive expertise and dreams. In D. Barrett & P. McNamara (Eds.), *Encyclopedia of sleep and dreams* (pp. 135–139). Santa Barbara, CA: Greenwood.

Kahan, T. L., & Claudatos, S. (2016). Phenomenological features of dreams: Results from dream log studies using the Subjective Experiences Rating Scale (SERS). *Consciousness and Cognition, 41,* 159–176. http://dx.doi.org/10.1016/j.concog.2016.02.007

Kahan, T. L., & Horton, C. L. (2012). Methodological challenges in the scientific study of dreams. In D. Barrett & P. McNamara (Eds.), *Encyclopedia of sleep and dreams* (pp. 418–421). Santa Barbara, CA: Greenwood.

Kahan, T. L., & LaBerge, S. (1996). Cognition and metacognition in dreaming and waking: Comparisons of first and third-person ratings. *Dreaming, 6*(4), 235–249. doi:10.1037/h0094459

Kahan, T. L., & LaBerge, S. P. (2011). Dreaming and waking: Similarities and differences revisited. *Consciousness and Cognition, 20,* 494–514. doi:10.1016/j.concog.2010.09.002

Kahn, D., Pace-Schott, E., & Hobson, A. J. (2002). Emotion and cognition: Feeling and character identification in dreaming. *Consciousness and Cognition, 11,* 34–50. doi:10.1006/ccog.2001.0537

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *American Journal of Psychology, 120,* 263–286. http://www.jstor.org/stable/20445398

Kramer, M. (1991). The nightmare: A failure in dream function. *Dreaming, 1*(4), 227–285.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Larsen, R. J., & Fredrickson, B. L. (1999). Measurement issues in emotion research. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 40–60). New York, NY: Russell Sage.

Link, K. E., & Kreuz, R. J. (2005). Do men and women differ in their use of nonliteral language when they talk about emotions? In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 153–181). Mahwah, NJ: Erlbaum.

Merritt, J. M., Stickgold, R., Pace-Schott, E., Williams, J., & Hobson, A. J. (1994). Emotion profiles in the dreams of men and women. *Consciousness and Cognition, 3,* 46–60. doi:10.1006/ccog.1994.1004

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods, 37,* 626–630. doi:10.3758/BF03192732

Miron-Shatz, T. (2009). Evaluating multiepisode events: Boundary conditions for the peak-end rule. *Emotion, 9,* 206–213. doi:10.1037/a0015295

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of yesterday's emotions: Does the valence of experience affect the memory–experience gap? *Emotion, 9,* 885–891. doi:10.1037/a0017823

Nielsen, L., & Kaszniak, A. W. (2007). Conceptual, theoretical, and methodological issues in inferring subjective emotion experience: Recommendations for researchers. In J. J. B. Allen & J. Coan (Eds.), *The handbook of emotion elicitation and assessment* (pp. 361–375). New York, NY: Oxford University Press.

Nielsen, T. A., Deslauriers, D., & Baylor, G. W. (1991). Emotions in dream and waking event reports. *Dreaming, 1*(4), 287–300. http://dx.doi.org/10.1037/h0094340

Pagel, J. F., Blagrove, M., Levin, R., States, B., Stickgold, B., & White, S. (2001). Definitions of dream: A paradigm for comparing field descriptive specific studies of dream. *Dreaming, 11*(4), 195–202. doi:10.1023/A:1012240307661

Parke, A. R., & Horton, C. L. (2009). A re-examination of the interference hypothesis on dream recall and dream salience. *International Journal of Dream Research, 2*(2), 60–69.

Revonsuo, A. (2000). The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming. *Behavioral and Brain Sciences, 23,* 877–901. doi:10.1017/S0140525X00004015

Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon*. Cambridge, MA: MIT Press.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44,* 695–729. doi:10.1177/0539018405058216

Schredl, M. (2002). Questionnaires and diaries as research instruments in dream research: methodological issues. *Dreaming, 12*(1), 17–26. doi:10.1023/A:1013890421674

Schredl, M. (2010a). Characteristics and contents of dreams. *International Review of Neurobiology, 92,* 135–154. doi:10.1016/s0074-7742(10)92007-2

Schredl, M. (2010b). Dream content analysis: Basic principles. *International Journal of Dream Research, 3*(1), 65–73. http://dx.doi.org/10.11588/ijodr.2010.1.474

Schredl, M., & Doll, E. (1998). Emotions in diary dreams. *Consciousness and Cognition, 7,* 634–646. http://dx.doi.org/10.1006/ccog.1998.0356

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52,* 591–611. doi:10.1093/biomet/1052.1093-1094.1591

Sikka, P., Valli, K., Virta, T., & Revonsuo, A. (2014). I know how you felt last night, or do I? Self- and external ratings of emotions in REM sleep dreams. *Consciousness and Cognition, 25,* 51–66. http://dx.doi.org/10.1016/j.concog.2014.01.011

Snyder, F. (1970). The phenomenology of dreaming. In L. Madow & L. H. Snow (Eds.), *The psychodynamic impli-cations of the physiological studies on dreams* (pp. 124–151). Springfield, IL: Charles S. Thomas.

Stickgold, R. (2017). Introduction. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (6th ed., pp. 506–508). Philadelphia, PA: Elsevier.

St-Onge, M., Lortie-Lussier, M., Mercier, P., Grenier, J., & De Koninck, J. (2005). Emotions in the diary and REM dreams of young and late adulthood women and their relation to life satisfaction. *Dreaming, 15,* 116–128. http://dx.doi.org/10.1037/1053-0797.15.2.116

Strauch, I., & Meier, B. (1996). *In search of dreams: Results of experimental dream research*. Albany, NY: SUNY.

Suslow, T., Battacchi, M. W., & Renna, M. (1996). The Italian version of the affective Gottschalk–Gleser content analysis scales: A step toward concurrent validation. *European Journal of Psychological Assessment, 12*(1), 43–52. doi:10.1027/1015-5759.12.1.43

Tov, W., Ng, K. L., Lin, H., & Qiu, L. (2013). Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment, 25*(4), 1069–1078. doi:10.1037/a0033007

Watson, D. (2000). *Mood and temperament*. New York, NY: Guilford.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. Ames: The University of Iowa.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Windt, J. M. (2013). Reporting dream experience: Why (not) to be skeptical about dream reports. *Frontiers in Human Neuroscience, 7*. doi:10.3389/fnhum.2013.00708

Windt, J. M. (2015). *Dreaming. A conceptual framework for philosophy of mind and empirical research*. Cambridge, MA: MIT Press.

Winget, C., & Kramer, M. (1979). *Dimensions of dreams*. Gainesville: University of Florida Press.

Yu, C. K.-C. (2007). Emotions before, during, and after dreaming sleep. *Dreaming, 17*(2), 73–86. doi:10.1037/1053-0797.17.2.73

Zadra, A., & Domhoff, G. W. (2017). Dream content: Quantitative findings. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (6th ed., pp. 515–522). Philadelphia, PA: Elsevier.

Zadra, A., & Robert, G. (2012). Dream recall frequency: Impact of prospective measures and motivational factors. *Consciousness and Cognition, 21,* 1695–1702. http://dx.doi.org/10.1016/j.con