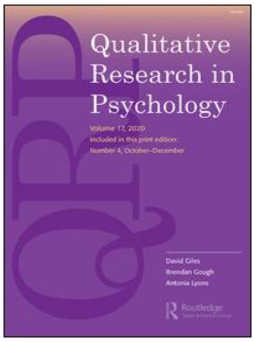

Routledge
Taylor & Francis Group

Qualitative Research in Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uqrp20

How questionnaires shape experienced symptoms. A qualitative case comparison study of questionnaire administration in psychotherapy research

Femke L. Truijens, Kimberly Van Nieuwenhove, Melissa M. De Smet, Mattias Desmet & Reitske Meganck

To cite this article: Femke L. Truijens, Kimberly Van Nieuwenhove, Melissa M. De Smet, Mattias Desmet & Reitske Meganck (2021): How questionnaires shape experienced symptoms. A qualitative case comparison study of questionnaire administration in psychotherapy research, Qualitative Research in Psychology, DOI: 10.1080/14780887.2021.1886383

To link to this article: https://doi.org/10.1080/14780887.2021.1886383

QUALITATIVE RESEARCH IN PSYCHOLOGY
https://doi.org/10.1080/14780887.2021.1886383

OPEN ACCESS

How questionnaires shape experienced symptoms. A qualitative case comparison study of questionnaire administration in psychotherapy research

Femke L. Truijens[a,b], Kimberly Van Nieuwenhove[b], Melissa M. De Smet[b], Mattias Desmet[b], and Reitske Meganck[b]

[a]Erasmus University Rotterdam, Erasmus School of Social and Behavioral Sciences, Department of Psychology, Education and Child Studies, Rotterdam, The Netherlands; [b], Faculty of Psychology and Educational Sciences, Department of Psychoanalysis and Clinical Consulting, Ghent University, Ghent, Belgium

ABSTRACT

Quantified symptom measurement by self-report questionnaires is part of the 'gold standard' of assessing psychotherapeutic efficacy. In this paper, we report a qualitative case comparison to explore how June and Amy, two patient-participants in a gold standard psychotherapy study, experienced the process of quantitative data collection. The study resembles cognitive interviewing studies conducted in the development of measures, yet advances them by investigating patients' experiences of questionnaire administration in actual psychotherapy. Both cases reported known issues in interpretation of pre-structured item- and response formats, communicative administrator-respondent dynamics, and response shifts. Beyond known scoring problems, the act of questionnaire administration changed their interpretation of experienced symptoms, which facilitated clinical change beyond therapeutic effects. For Amy, this change was associated with improvement, but for June, questionnaire administration facilitated deterioration in experienced symptoms. These findings emphasize that it is both epistemically and ethically vital to consider measurement effects in clinical practice. This study demonstrates the importance of taking a qualitative stance in psychotherapy research, as qualitative research can elaborate the contextual and idiosyncratic nature of questionnaire scores, and highlights that both researchers and clinicians have to be attentive to the meaning of scores as words in participants' clinical stories.

KEYWORDS
Questionnaire experience; patient-Reported Outcome Measures (PROMs); measurement effects; psychotherapy Research; response Shift; qualitative Research

Introduction

Psychotherapy research forms the scientific foundation of 'Evidence-Based Treatment' and is known for its emphasis on the use of systematic data collection and rigorous analyses in order to formulate sound conclusions on the efficacy of treatments (Chambless and Ollendick 2001). The 'gold

CONTACT Femke L. Truijens truijens@essb.eur.nl Erasmus University Rotterdam, Mandeville Building, Room T13-20, Thomas Morelaan, 3062 PA Rotterdam, The Netherlands.

Supplemental data can be accessed here.

standard' research procedure is based on quantitative assessments of symptoms before and after treatment, to deduce general treatment efficacy from 'the data' (McLeod 2001; Wampold and Imel 2015). The measures used to assess such data are usually self-report questionnaires (i.e., patient-reported outcome measures, PROMs). PROMs are often preferred over other types of observation because they are standardized: pre-formulated symptom items are scored quantitatively resulting in uniform type of data that can be aggregated over people and statistically analysed to derive general evidence (McLeod 2001). In this paper, we take a qualitative stance (Hesse-Biber 2010) to explore how patient-participants in psychotherapy research experience the *process of questionnaire administration*, to understand how their experiences are translated into 'the data' that will form the evidence-*base* in psychotherapy research.

This paper introduces June and Amy,[1] two patient-participants who participated in The Ghent Psychotherapy Study (GPS, Meganck et al. 2017). June and Amy both applied for participation in the GPS with an overwhelming fear of being 'mentally ill' or – in their own words – 'crazy'. In the general GPS sample, June stood out because of her tendency to annotate her paper-and-pencil questionnaires with an array of written and visual remarks. In our previous qualitative analysis of June's annotations (Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019), we found that the *act* of questionnaire administration had a substantial impact on the way in which June experienced her presenting problems. June has often been called an outlier,[2] because of her notable difficulties with scoring the questionnaires. However, as the experience of questionnaire administration is seldomly studied systematically in a clinical context, it was also hypothesized that her experiences may be more common for patients with anxiety issues. The rationale for the current study was to perform a case comparison study based on a second in-depth analysis of questionnaire administration for a patient with similar anxiety symptoms. The case comparison design allows for analysis of similarities *and* divergences (Flyvbjerg 2006) in the experiences of questionnaire administration, to further understand and elaborate on how the process of questionnaire administration could shape 'the data' that is used for inference of general treatment efficacy.

To secure the soundness of data as input for the inference of treatment efficacy, psychotherapy researchers generally rely on validated measures. When measures are validated, this means that 'it measures what it purports to measure' (Borsboom, Mellenbergh, and Van Heerden 2004, 1061). Provided that respondents interpret items as intended by the questionnaire developer and score honestly and accurately given their experience and situation, the validation of the measure implies that resulting measurement data indeed captures the intended construct accurately (Truijens et al. 2019a). To ensure that respondents indeed interpret items as intended, the development of a measure yields a phase of cognitive interviewing, in which proposed respondents are asked to ‘think aloud’ while scoring items (Peterson, Peterson, and Gilmore Powell 2017). In this validation process, the goal is to identify items that are incomprehensible, selective or multi-interpretable, for example.

Once a measure is validated, test *users* tend to take the validation as a warrant that scoring issues are covered, such that ‘the data’ can straightforwardly be used for diagnostic and epistemic purposes (Moss 2013). However, the interpretation process is not complete once the development of the measure is finished. Asking respondents to think aloud helps trace divergences from the intended interpretations, yet ‘cognitive issues’ may be more subtle than can be traced in cognitive interviews, as respondents may not be aware of issues that affect their scoring (Schwarz 1999). For example, answering styles can be primed by item order, items can be stereotyped or culturally biased, and scoring options can be shaped by the format of response scales that involve extreme scoring options, which are less likely scored than less extreme scoring options. Moreover, questionnaire administration may involve subtle ‘communicative’ factors that can influence scoring (Schwarz 1999). For example, the number of times a certain symptom is assessed in a questionnaire may suggest that those items are more important than others.

Communicative issues go beyond psychometric features of the measure, as they can occur *after* the development of the questionnaire and are therefore not covered by the validation process. In their cognitive interviewing Galasiński and Kozłowska (2010, 2013) found that respondents tend to take questionnaire administration as a way of interacting with the questionnaire administrator: they take the interest or purpose of questionnaire administration into account, and answer the items accordingly. That way, scores become a dialectical form of communication rather than a neutral or context-independent report of experienced symptoms.

Importantly, in psychotherapy research, measurement is not just conducted to assess symptom levels as such. Rather, data are collected to calculate *symptom change* over the course of treatment. Subsequently, in gold standard psychotherapy research, pre-post differences are interpreted as *effect* of a psychotherapeutic intervention – provided that they were found under controlled circumstances (Chambless and Ollendick 2001). As Schwartz and Rapkin (2004) warn, however, assessment of change is prone to the phenomenon ‘response shift’. Following an intervention, assessed symptoms may be reduced when the symptom is in fact experienced in a less severe manner, but they also be *interpreted differently*. For example, a patient may not be able to walk more meters after a hip surgery, but *does* experience improvement because the same amount of walked meters is experienced as a stability rather than a sign of decrease. This response shift threats the validity of pre-post differences, as similar numbers *could* mask change in experience. Response shifts may occur as recalibration (e.g., recognizing that the initial score was actually too low), reprioritization (e.g., experiencing similar symptom levels but interpreting them as less important), and reconceptualization (e.g., understanding that symptoms should be interpreted differently, thus implying a different score). Westerman et al.'s (2008) study indicated frequent occurrence of response shifts when there were *repeated* assessments.

These studies emphasize that meaning-making processes in questionnaire administration are *contextual*, that is, embedded in the concrete situation of administration in which respondents make *sense* of items in questionnaires in a cognitive (1), communicative (2) and sometimes changed (3) way of interpretation. The current study advances cognitive interview studies by scrutinizing experiences of questionnaire administration in the concrete context of 'gold standard' psychotherapy research. In this, we do not regard meaning-making as 'covered' by validation nor as finished once the measure is developed. Rather, we explored meaning-making processes involved within the actual context of questionnaire administration to assess change over the course of psychotherapeutic treatment.

Specifically, we explored how self-report questionnaires were experienced by our two cases, June and Amy. In the GPS, patients' symptoms were assessed by a battery of 'gold standard' validated PROMs (Meganck et al. 2017). To enable qualitative analysis of treatment processes, the standard pre-post design was complemented with cognitive interviews on how the two patient-participants experienced symptom change and therapeutic processes. For the purpose of validation of treatment effects, the interview closed with an open question about how they experienced the research procedure and the therapy-research relationship. Beyond the interviews, all therapy sessions were audiotaped, which allowed us to also analyse spontaneous rather than asked-upon experiences of questionnaires administration. As Galasiński and Kozłowska (2010) note, the limitation of straightforward cognitive interviewing techniques is that they may prime, probe or frame experiences in a dialectal interaction with the researcher. This stresses the importance of spontaneous speech, and therefore we used *all* narrative data provided by both cases, both within the interviews and s throughout therapy. In the discussion, we elaborate how the findings go beyond known measurement problems, and we argue that these issues are especially salient in the context of psychotherapy research, where measurement can both affect epistemic and ethical dimensions of clinical research and practice.

Method

Participants

Case selection

The cases were selected on the similarity in their presenting problems, as they both expressed severe anxiety that they might be 'mad' or 'crazy' (see findings) associated with experienced feelings of depression and hopelessness, which made them eligible for participation in the depression treatment study (Meganck et al. 2017). Both cases have been previously studied: Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck (2019) studied the function of June's annotations on the paper-and-pencil questionnaires in her meaning-making process; Van Nieuwenhove et al. (2019) studied Amy's case with focus on therapeutic interventions in case of complex trauma. *The cases were purposively.[3] selected for the current follow-up case comparison, based on the similarity of their *experienced symptoms.[4] at the start of treatment. June can be considered an extreme case (Flyvbjerg 2006), given her explicit questionnaire annotations (Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019), but Amy is a rather typical case (Willig 2012) who did not stand out in the GPS sample other than her resemblance with June's core presenting problems.

June

At the start of treatment, June was a 26-year-old Caucasian female, who applied for therapy upon referral by her general physician. During the intake phase of treatment, she was asked to join the pilot phase of the psychotherapy study. When June applied, her presenting problem was an intense and overwhelming fear of being sick, dysfunctional or abnormal. Her therapist was a 36-year-old Caucasian male, with 6 years of clinical experience in Psychoanalytic Psychotherapy at the start of treatment. In total, she had 53 sessions of supportive-expressive therapy (cf., Luborsky 1984).

Amy

At the start of treatment, Amy was 26-year-old Caucasian female, who applied for therapy after the recruitment campaign of the GPS. Upon application, she experienced overwhelming moments of anxiety and emotional outbursts, which she took as a strong sign of mental disease, abnormality or craziness. Furthermore, she experienced depressive symptoms that were diagnosed as Major Depressive Disorder (see procedure). Her therapist was a 30-year-old Caucasian female, who had 7 years of clinical experience in Psychoanalytic Psychotherapy at the start of treatment. In total, Amy had 20 sessions of Supportive-Expressive Therapy (cf., Luborsky 1984).

Procedure

Data collection

This case comparison study is based on a selection of data collected in GPS (Meganck et al. 2017). June participated in the pilot phase, in which baseline assessment was conducted using a test battery of validated self-report measures – including the Symptom Checklist (SCL-90; Derogatis 1992), the Beck

Depression Inventory (BDI-II-NL; Beck, Steer, and Brown 1996), and the General Health Questionnaire (Koeter and Ormel 1991) – a number of general well-being- and personality measures, and a clinical diagnostic assessment based on DSM-IV criteria. Amy participated in the randomized controlled trial (RCT), in which baseline assessment yielded a similar battery of validated self-report measures, a structured diagnostic interview (SCID-I & II; First et al. 2002) and a semi-structured clinical assessment (CDI; Westen 2002). Furthermore, patient-participants were asked to formulate one to five idiosyncratic items, in which they worded and scored their most prominent experienced symptom(s).

During the course of treatment, all therapeutic sessions were audio recorded. Each therapy session was accompanied by a small test battery (every session) or a large test battery (every fourth session). In the pilot phase, patient-participants were asked to participate in four follow-up interviews over the course of two years. In the RCT, patient-participants were asked to participate in a peri-interview, a post-interview, and four follow-up interviews over the course of two years. The interviews were semi-structured Client Change Interviews (Elliott 1999) and were audio recorded. All narrative data was transcribed verbatim, and translated from Dutch to English for the current paper.

Data analysis & quality control

June was previously studied (Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019) in two rounds of coding that were focused on experienced symptoms and experience of questionnaire administration. Amy was previously studied in a single-case study with a different focus (Van Nieuwenhove et al. 2019). For the current study we re-analysed this case for the purpose of systematic theme-based comparison. In the first phase, the first and second author both identified themes that were related to Amy's experienced symptoms (consensus-based, see Hill, Chui, and Baumann 2013). In the second phase, Amy's experiences of questionnaire administration were analysed in all interviews (asked-upon narrative) and therapeutic narratives (spontaneous narrative). Data excerpts and identified themes were iteratively checked by the first and second author (cf., McLeod and Elliott 2011).

In the third phase, the main findings with regard to symptom experience and experienced questionnaire administration for both cases were compared to derive higher order themes (Willig 2012). After forming the higher order themes, they were iteratively grounded in the case-specific findings to check the fit of themes with each case narrative (Stiles 1993). Finally, the findings of the third round of analysis were written up by the first author, and interpretations were discussed in-depth with the fourth author (Creswell and Miller 2000).

Throughout the analytic procedure, we undertook a number of strategies to ensure quality control. To avoid confirmation bias (Flyvbjerg 2006), we first analysed each case independently, with different research teams. Subsequently, the cases were compared by the first author (full familiarity with both cases and current research questions), the second author (full familiarity with both cases in different research context, no previous familiarity with research questions) and the third author (no previous familiarity with cases nor research questions) to balance the knowledgeability in the audit team. The themes in the comparison were systematically cross-checked by members of the research teams of Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck (2019) and Van Nieuwenhove et al. (2019), who knew the cases from another research angle and thus could balance prioritizing and interpretation of themes (Creswell and Miller 2000).

Findings

In the first section, we discuss main themes in the presenting problems expressed by Amy and June (see online appendices 1 and 2 for main themes in the individual cases). In the second section, we discuss four higher order themes in the experience of questionnaire administration that were found by comparing Amy's and June's experiences.

Presenting problems as experienced by June and Amy

When Amy applied for therapy in the context of the research, her presenting problem was the overwhelming fear of being crazy or abnormal, that is, deviating from societal norms. She experienced regular episodes of overwhelming and surprising emotions, and she suffered from black-outs, nightmares and – for her – irrational behaviour. After treatment termination, Amy identified these emotions and behaviours as related to complex childhood trauma (cf., Van Nieuwenhove et al. 2019), but at intake, she called them 'a little monster within me' (Intake-interview) and regarded them as signs of craziness.

> 'Like, not trusting my emotions. [...] I needed constant confirmation [...] because I thought it was a sign that I was totally crazy.' (Amy; Post-interview)

Because the emotions constantly surprised her, she feared that she could not *trust* her own thoughts and emotions, and that her memories were made-up – which she had been accused of often in the past by people whom she trusted. In the past, family members denied that situations that she remembered did happen. Therefore, being overwhelmed by such memories and experiencing anxiety and emotional suffering, did feel like signs of craziness for Amy.

> 'I feel like I have less and less control over my own thoughts. Sometimes a thought pops up in my head, of which I really don't know if I truly feel that or not. [...] so I am troubled in knowing how I actually feel about something.' (Amy; Intake-interview)

Amy's therapeutic aim was to get help identifying which thoughts are 'right' and 'wrong'. She formulated this question in her idiosyncratic items as 'the question which emotion is correct' and 'fear that I cannot trust my emotions'.

These questions were embedded in her fear that thoughts, desires, and actions may have 'consequences' in the eyes of others. Because she was so afraid that others would observe flawed thoughts and behaviours, and interpret these as signs of craziness, she was constantly scared to make decisions:

> '[Like] putting the potato peeler on a wrong place and then thinking that I'm going completely insane'. (Amy; Post-interview)

Her anxiety to make wrong decisions was connected to her fear that *others* would lock her up for being crazy; in fact, Amy's father often threatened to put her in 'the crazy house' near their home (Amy; Intake-interview). To prevent others from observing signs of craziness, Amy had an obsessive tendency to control herself, by being overly rational and judgemental towards herself.

> 'I do not have absolute knowledge about myself. That frustrates me, because I am very analytical [...] I am afraid to lose control. [...] And then I suddenly realise that and I think: "everybody can see that and will think that I am completely insane".' (Amy; Intake-interview)

As Amy constantly scanned her thoughts, she found it difficult to answer emotion-focused questions and searched for nuance, detail and factual self-awareness. In this way, she was constantly looking for a norm for her own behaviour and thoughts:

> 'It should not affect me. I feel that I should be able to. to respond to it in an objective manner.' (Amy; Intake-interview)

Reiterating, in Amy's therapeutic narrative, the fear of being crazy was embedded in a fear that others would *find* her crazy, which gave 'the other' a prominent role in her experienced anxiety, as well as in her attempted behaviour towards others.

June also started treatment with an overwhelming fear of being crazy, ill or abnormal. She was obsessed with being able to suppress or hide possible craziness before others could observe it. Consequently, June was constantly searching for 'signs' in her own behaviour, feelings and thoughts:

> 'So I want to stay in bed and leave everything alone. And then I think: see, it is a sign, I am depressed! When people want to stay in bed .... That is a sign right ... ' (June; Session 29)

Indeed, June read particular behaviours *as* signs of the craziness that she fears. Importantly,, June was highly susceptible to suggestion. Her anxiety could peek by suggestions from all kinds of sources, ranging from newspapers items and overheard conversations, to her own thoughts and behaviours:

> ‘Like, I was afraid for a lot of things again. Yes, then I think that I have a **depression** again [. . .] It was on a television show a couple of weeks ago, I think. Something about women with **substance abuse disorder** . . . And then I freak out again [. . .] And then I think, like, “I have that myself”. And then I think “I have **paranoia**, or what is this?” Yes, then I become super afraid, or like “I am depressed” or like, agitated, like . . . And then I think like, “I *must* be depressed”.’ (June; Session 6; emphasis added)

In this quote, June mentions three apparently random syndromes within the span of a minute, showing that she did not necessarily have to experience concrete symptoms to still fear the presence of a particular syndrome. Indeed, the suggestion in a television show makes her fear that she *can* have those particular symptoms – even though she does not experience any of them – and as a consequence she obsessively scans her own behaviour and thoughts to be sure that she can identify them before others will.

Ultimately, June feared that if others did perceive her as crazy, they would lock her up in a psychiatric facility and exclude her from society.

> ‘Because it really is that fear for me, that the people will abandon me, that they will not see me anymore . . . Like when I would have [a] depression, that people would put me in a crazy house [. . . and] would not visit me or that they would not talk to me anymore.’ (June; Session 40)

Given her fear of the judgement of others, June constantly searched for *norms* to ensure that her behaviour fell within the normal range. Importantly, she was not quickly eased, as she kept doubting reassurances from the therapist and looking for counterevidence in her own thinking.

Table 1. Synthesized themes in the experience of questionnaire administration by June and Amy.

Questionnaires do not allow for nuanced and straightforward reports
Complex meaning in items
Complex instructions and time indications in questionnaires
Questionnaire scoring does not allow for nuanced answers
Trouble to quantify experiences
Questionnaires decontextualize symptoms
Questionnaire administration forces self-evaluation and self-awareness
Questionnaire administration forces reflection on own complaints
Repeated self-monitoring and self-evaluation
Understanding apparently random or independent behavior as ‘signs’
Questionnaires suggest a norm for (ab)normality of behaviour
Questionnaires suggest a norm on normal and abnormal behavior
Questionnaires allow others to deduce ‘profile’ of craziness
Questionnaire administration impacts the interpretation of core complaints
Questionnaire administration impacts the interpretation of complaints as ‘signs’
Questionnaire can provoke a nuanced interpretation of craziness
Questionnaires can provoke priming and substantiation of craziness

Bold text = main themes, regular text = subthemes, *italic text* = subthemes.

Experience of questionnaire administration

In this section, we discuss four higher order themes in the experience of questionnaire administration that were identified in the comparative analysis of Amy's and June's experiences (Table 1).

Questionnaires do not allow for nuanced and straightforward reports

Both June and Amy reported finding it hard to fit their genuine experiences to the pre-formulated items, the pre-structured scoring format and the instructions regarding time and references:

> 'Then I was so conflicted about it all, when it was asked in such a hard way. Like with time. or. like how something manifests, like, "is it this way or that way". Then I'm not sure: I have this, so does that mean a "yes" or a "no".?' (Amy; 6-month Follow-up)

> 'So, like, uhm . . .I have to answer here, at "in general," what does that mean, "the last three weeks"? Because the last three, four weeks it is not good, but. What does "in general" mean then?' (June; Session 57)

Furthermore, Amy reported that questionnaires assess experienced symptoms in a 'decontextualized' way (Amy; Peri-interview), which did not allow her to provide nuanced answers. June similarly experienced the questionnaires as too limited to report her experiences with sufficient nuance, and she felt troubled by quantifying her experiences accurately:

> 'But when I notice that I am angry or anxious, I like cannot really categorize that or give a value to it. What number is that then, that anger?' (June; Session 58)

June therefore often scored multiple options: to show that she scores in between two numbers, she circles two numbers on a scale and connects them, or she scores one number with a thick line and scores a second number with a thin line or between brackets (Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019, Figure 4). Furthermore, to address the lack of nuance in questionnaire response options, June created extra scoring scales for different references. For example, she scored items differently for 'professional' and 'personal' circumstances (June; Session 57), or for 'people that are "close" → parents, partner' and 'friends and third parties' (June; Session 13).

So while both Amy and June were motivated to score the questionnaires as accurately as possible, they were both worried that the pre-structured format of items and responses did not allow for accurate and nuanced reports of their experiences.

Questionnaire administration forces self-evaluation and self-awareness

For both June and Amy, the questionnaire administration ‘forces’ (Amy, Post-interview) to evaluate themselves. According to Amy, scoring questionnaires routinely increased her awareness of the presence and development of symptoms which initiated a continuous process of self-reflection and evaluation of symptoms. Moreover, certain items made her aware of behaviours and thoughts that she did experience but did not concretely notice before:

> ‘Specifically, in [those questionnaires . . .], there are very specific questions that I would never have thought about otherwise, like “aha, that’s something that I do or that other people do, I did not know that”.’ (Amy; Post-interview)
>
> ‘The questions are so odd sometimes, which I never thought about before, and then I realize that is actually present. Like, the feeling of guilt, I notice that way more because of those questionnaires. [. . .] that’s something that I would not have noticed otherwise, that I actually experience that very often.’ (Amy; Peri-interview)

Questionnaire items thus suggest symptoms that Amy did not notice before, and lead to her observing her behaviour and noticing regularities. As such, the act of questionnaire administration initiates self-monitoring beyond initial experiences. Moreover, the regular scoring became a tool for Amy to distinguish and prioritize certain feelings over others. Amy noticed that the instructed time frame is actually *helpful* for her:

> ‘Yeah, the questionnaires help me with checking [. . .] how I feel in general and how I feel within a week [. . .]. I find that helpful, because I very quickly feel that a specific mood, a negative mood [. . .] is persistent, but then I realize, no, I only had a bad week, I did have one severe anxiety attack, but that does not mean that I am panicking all the time.’ (Amy; Peri-interview)

The fact that Amy is ‘forced’ (Post-interview) to reflect on her experienced symptoms in short time frames in some questionnaires, and on general tendencies in others, helped her to distinguish between general and time-specific behaviour. This provided Amy with a new perspective on her experienced symptoms: whereas she used to explain her experiences as ‘signs’ of general craziness, monitoring time-specific behaviour allowed her to disconnect her fears from actual behaviour. This way, the questionnaires served as an intervention to the interpretation of her own behaviour, which started a process of reinterpretation of her fears.

Likewise, June experienced questionnaires as suggesting particular symptoms that she was not aware of before. Importantly, however, for June this suggestion was rather problematic, as she was highly susceptible for all kinds of suggestions that spark her fear. As June feared that she may show signs of craziness that others could perceive before she was able to eliminate them, she was constantly screening all kinds of sources (from newspapers to concrete behaviour or particular thoughts):

> 'Like, sui-, the word suicide comes up in my mind and then I am extremely afraid, I don't want that [*breathes heavily* . . .]. I am here by myself so in theory I could. dó something like that, there is no one here to control me.' (June; Session 25)

While June did not experience suicidal tendencies, the confrontation with the *concept* of suicidal ideations triggers her anxiety, showing how the questionnaires became suggestive to June and triggered her fears. She explicitly emphasized this affect by annotating the suicide items time and time again, for example with the remarks "scares me, want it to go away!!!" (BDI item 9, session 56) and "from this thought, I get scared" (SCL item 15, session 56)(see further Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019).

Regularly completing questionnaires set in motion a process of self-observation and reflection on experienced symptoms both for Amy and June. This self-evaluation was not limited to awareness of symptoms that they actively experienced, as it also yielded suggested symptoms that they were not (so) aware of previously. As such, the questionnaire administration meant they experienced thoughts and behaviours that they would not have otherwise. For Amy, this has a positive effect: it became a way of putting actual experiences in perspective. For June, in contrast, suggested items on the questionnaire fuelled her anxiety and the questionnaire administration became a catalyst of her fears.

Questionnaires suggest a norm for (ab)normality of behaviour

June and Amy experienced the questionnaires as implying a norm of (ab) normality. Both June and Amy were strongly focused on norms to judge whether they could in fact be crazy. As the questionnaires cover a whole range of possible abnormal thoughts and behaviours, the questionnaires start to serve as such a norm.

Whereas the previous theme showed how the questionnaires suggested possible symptoms to June and Amy, for Amy it had an opposite effect as well: the questionnaires made her aware of symptoms that other people might experience, but which she does *not*. This way, she comes to understand that other people may experience different and – importantly – worse symptoms than the 'symptoms' that she experienced herself.

> 'Like, number one of confronting items are those suicide-items. [. . .] That was something that I have thought about whole my life [. . .] I think of suicide, but I would never do it [. . .] And I thought that was so problematic. But now I feel like [. . .] okay, this is something that people just can have.' (Amy; 6-month Follow-up)

In fact, Amy had experienced suicidal ideations throughout her life. In the administered questionnaires, suicidal ideations are assessed multiple times. In the Beck Depression Inventory (Beck, Steer, and Brown 1996), for example, question 9 is captioned “suicidal thoughts or wishes”, and it has to be scored on a scale that provides the following labels: 0 = ‘I don’t have thoughts of killing myself’, 1 = ‘I have thoughts of killing myself, but I would not carry them out’, 2 = ‘I would like to kill myself’, and 3 = ‘I would kill myself if I had the chance’. As Amy has always had thoughts about suicide but would not pursue it, she continuously scored 1 out of 3. Consequently, the scale helped Amy to reframe the suicidal ideations that she always interpreted as a strong sign of craziness as actually being relatively mild.

This way, for Amy, the questionnaires became normative: they showed her how she could interpret her experienced symptoms in relation to others, which leads to a more nuanced understanding of herself. Whereas she tended to interpret her symptoms as ‘signs of craziness’, her awareness of the possible nature and severity of symptoms for others put her fear in perspective and reduced her anxiety:

> ‘Like, becoming aware of that I do that, but also on the other hand the feeling that, like: “okay, so that is something normal, that is something that can just happen every once in a while.’ (Amy; Post-interview)

For June, however, the normativity in questionnaires is anything but soothing. She feared that when others viewed her scores, they may deduce a ‘profile’ (Post-interview) or image of her that verified her craziness. More specifically, as the questionnaire items explicitly highlight content on which she *could* be judged, completing the questionnaire manifested the judgement of ‘the other’. Therefore, June tended to be as nuanced and detailed as possible in her questionnaire scoring, to ensure that the administrator or reader would understand her intended interpretations. By annotating and clarifying her scores, she addressed the other directly, as a reader who intends to make conclusions about her mental status.

June was very anxious that such a verdict might lead to exclusion and lock-up, so for her the questionnaire is a manifest set of normative items that suggest the kinds of behaviour that she has to avoid or control to stay within the ‘normal’ range:

> ‘So what number is that, then, that anger? [. . .] And is that acceptable? And, does that fall under the label normal? Like, is that normal, and to what extent . . .’ (June; Session 58)

Similarly, Amy was apprehensive about the idea that others might deduce a ‘profile’ from her scores that might indicate the craziness that she always feared.

> ‘I immediately tried to see how it would be analysed, what it would mean. In what clinical scheme I would fit [. . .] like I imagine how a therapist [looking at] the questionnaires and saying: “oh, wow, this is a classic case of . . .”’ (Amy; 6-month Follow-up)

However, through the repetition of questionnaire administration and the nuance that was forming in her interpretation of self, Amy's image of other's expertise starts to change:

> 'There are a lot of people who wanted to give me a bit like, a diagnosis. From their own [. . .] "Wikipedia-expertise". [. . .] But I found [that those questionnaires . . .] helped me so much. And that's funny because those questions are more open now. That I, that I don't have to write myself off anymore because I scored a two or a five" (Amy; 6-month Follow-up)

The questionnaire administration thus not only allowed Amy to perceive her experienced symptoms in a more nuanced way, she also put the 'expertise' of others in perspective. The verdict of others played a prominent role in her fear, so the fact that she started to doubt the expertise of others had a large impact on her experienced fear. For Amy, the questionnaires suggest a norm on (ab)normality of behaviour, a point of reference from which she could test her craziness, and consequently, she had less need for others as a reference of normality. For Amy, the questionnaires thus made others less scary.

For both June and Amy, the questionnaires functioned as a tangible 'norm' on normal and abnormal behaviour. For Amy, this had a positive effect as it allowed her to nuance her experienced symptoms and to distance herself from the judgement of others. For June, however, the fact that questionnaires will be viewed by an actual reader made the judging other manifest, which increased her fears (Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019). The confrontation with norms on (ab)normality of behaviour thus had a substantial impact on both patients, albeit in opposite directions.

Questionnaire administration impacts the interpretation of experienced symptoms

Given the perceived normativity of the questionnaires, the act of questionnaire administration actively impacted both June's and Amy's interpretation of presenting problems. For Amy, the questionnaire administration became an external point of reference (Peri-interview), and she explicitly distinguished this questionnaire effect from therapy.

> 'I like it that that there is, sort of, an entity outside of the world of the [therapeutic] sessions. [. . .] that is more like the real world, [rather than] the internal therapy world.' (Amy; Peri-interview)

This external point of reference became a new standard for Amy to test her own fears as well as others' opinions. This way, she learnt that she can trust her own thoughts more than others'. One of her idiosyncratic items yielded her fear that she could not trust her own emotions and thoughts. Consequently, she constantly consulted others to verify what was correct or crazy, but that also made her fear others as evaluating and normative. However, when the questionnaires started functioning as an external norm, she no longer needed other people's 'Wikipedia-expertise' (6-month Follow-up). So, the questionnaire not only made her re-interpret her experienced symptoms as signs of craziness, she also re-interpreted others as important reference points.

This change is not simply a reduction of Amy's experienced symptom, but it involves a reconceptualization of her fear, which is related to a reconceptualization of herself in relation to others. This makes the evaluation of others less threatening, and therefore results in a more nuanced interpretation of her own fear. That does not necessarily imply that she does not experience the overwhelming emotions anymore (i.e., symptom reduction) but it implies that her *interpretation* of these overwhelming emotions was changed. Thus, as the questionnaires started to function as an external reference point for Amy, the main features of her experienced symptoms changed.

For June, the questionnaires also affected the main features of her experienced symptoms, but in the opposite direction. A core feature of June's symptoms was her vulnerability to suggestions which she tended to take as signs of her own craziness, and tried to control by continuous screening of her own thoughts and behaviour. The questionnaires systematically confront her with an exhaustive set of possible symptoms, which make the suggested 'signs of craziness' *manifest* rather than random and incidental. Moreover, as the administrator can deduce a profile of her, the questionnaire administration makes the judging other concrete and present. Consequently, she can no longer use her preferred solution of avoiding or controlling others' perceptions by screening her own thoughts and behaviours first. Rather than controlling herself, she now also actively has to control the reader of the questionnaires. Indeed, she tried to control what the reader would take from her scores by explicitly addressing her 'audience' in annotations and by differentiating her scoring.

For June it is very important to be as honest as possible, but that also forces her to often score on the higher ends of the scale. However, this high scoring is a concrete and visible confrontation with how her experienced symptoms compare to others'. This not only increases the severity of her fear (i.e., symptom increase), it also starts to function as a 'guarantee':

> 'It needs to be black-on-white. I said that before, I am such a person who is very visual. [...] I do not have any guarantees, [...] it's only when I would see it, that I would think like "yes, indeed!"' (June; Session 2)

The questionnaire items are visual and concrete such that they *become* guarantees for June. However, rather than solving fears, this guarantee makes the suggestions manifest and consequently limits the sense of control of her 'signs', even in situations where there is no precedent:

> ‘It is the next [item]: “To feel an urge to hit, hurt or injure others.” [. . .] That scares me so much that I am gonna be focused on that all week. [. . .] I’m gonna put that in my mind now, and that I will go, like, it will trigger “see, see, I have that”.’ (June; Session 57).

So for June, the questionnaires became an external source of fears, which increases both the severity and scope of feared diseases. Therefore, the questionnaires not only worsen June’s level of fear (i.e., symptom severity), but also actively change her way of *interpreting* the status of those fears.

Both for Amy and for June, these changes affected their interpretation of ‘signs’. Both feared that their thoughts, emotions and behaviours showed a variety of ‘signs’ of craziness which would motivate others to lock them up in a psychiatric facility. For Amy, the questionnaire as external reference made her understand that having a bad day does not signal complete craziness immediately. Therefore, the questionnaire administration allowed her to stop interpreting all kinds of thoughts and behaviours as ‘signs’. Moreover, she started to see that certain verdicts of others might rather have been signs of *their* imperfection (2-year Follow-up).

By changing her interpretation of emotions as ‘signs’ of disease, the core of Amy’s symptom was changed: while she still experienced overwhelming emotions, the changed interpretation prevented her from being so scared by them. Moreover, by the act of questionnaire completion, Amy found herself being capable of distinguishing herself from specific scores on specific items. She came to see that certain scores were bound to time-specific situations, reactions to situations, recurring childhood memories, et cetera. The act of questionnaire completion for Amy became an exercise to not see herself in ‘absolute’ terms (Intake-interview) nor to expect herself to reach a ‘perfect zero’ (6-month Follow-up), which led to more acceptation of herself as a person (Post-interview). This way, the act of questionnaire administration for Amy led to normalization and to overall ‘empowerment’ (Post-interview).

For June, it had the opposite effect: not only was she prompted in her search for signs, it also suggested new kinds of signs that catalysed her fear. Moreover, the fact that she had to score high on the scales, substantiated a realness of her fears, and as such it changed the nature of the fears, as they became ‘visible’ signs that have a ‘real’ presence for her. This way, the act of questionnaire administration for June led to priming of new ‘phobic stimuli’ (Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019), and as such substantiated June in her fear of being deviant.

So, both for June and Amy, the act of questionnaire administration impacted the interpretation of the symptoms that got them to seek treatment in the first place. For both patient-participants, this yielded an active *change* in their interpretation of experienced symptoms. In Amy’s case, she did experience the same overwhelming emotions, but no longer interpreted them as signs of overall craziness. This allowed her to put her perception of self and her fear for others into perspective, which made her re-evaluate her tendency for control over herself. In June's case, she did experience the same overwhelming fears, but as the questionnaires provided an exhaustive list of possible abnormalities, the feared illnesses became manifest rather than random suggestions. Moreover, the other as judgmental and excluding became a concrete person, which imposed a new fear to her that her regular control would no longer suffice.

Discussion

In psychotherapy research, the aim is to gather evidence on the efficacy of treatments. Efficacy is operationalized as symptom change over the course of treatment, captured by pre- and post-treatment data collected via self-report questionnaires. In this paper, we scrutinized how the process of data collection by means of such self-report questionnaires was experienced by June and Amy, two patient-participants in our psychotherapy study (Meganck et al., 2017). Despite validation, self-report measures are associated with scoring issues (1), administrator-respondent dynamics involved in questionnaire administration (2), and response shifts in assessment of *change* (3). These reported threats to validity of data have a common root in the *contextuality* of response behaviour. Beyond these merely methodological threats, the findings of this study indicate that questionnaire administration can *interfere with* the treatment process and the understanding of experienced symptoms (4). Importantly, this finding emphasizes that measurement in the context of psychotherapy research does not only have epistemic implications, but also becomes a risk to clinical ethics. Below, we discuss in detail how known psychometric problems in a clinical context can *change* the experienced symptoms under assessment, and we argue for taking a qualitative stance (Hesse-Biber 2010) to address these epistemic and ethical consequences.

First, the case comparison study emphasized that questionnaire administration is an interpretative process, in which respondents actively try to understand meaning purported by the measures' items. The first theme captured the trouble that the patient-participants reported with interpretation of items and instructions, complexity of quantifying experiences, and feeling puzzled on what or whom to take as a reference point. As Schwarz' (1999) notes, respondents have to make sense of the items as *intended* by the developers. Despite the use of cognitive interviewing in the development of measures, the first theme resonates with Schwarz' (1999) warnings that the *interpretation process* is not finished nor covered once development is completed (cf., Moss 2013).

Moreover, interpretation issues do not only occur when respondents try to interpret pre-structured items, but also the other way around as respondents try to score their experiences genuinely, which is vividly illustrated by June's tendency to annotate the questionnaires. Specifically, respondents struggle with the translation of their experienced symptoms into pre-structured item-and response formats. In this, the communicative features of questionnaires (subtly) suggest order, priority, and weight of certain experiences, which may impact the responses of respondents (Schwarz 1999). While quantitative measures are generally preferred because of their ability to *reduce* or decontextualize experiences to form comprehensible and comparable data, the first theme emphasizes the importance of the contextual embeddedness of questionnaire responses.

Secondly, the communicative character of questionnaires is not limited *within* the measure, but is embedded in a respondent-administrator dynamic. As the third theme makes clear, in scoring questionnaires, patient-participants actively address an *audience* (the administrator, the reader). As Amy and June were inclined to interpret their own experiences through the lens of the administrator, they experienced the questionnaires as *normative*. Similarly, Galasiński and Kozłowska (2013) found that respondents are 'are not merely trying to get the questionnaire completed but also to do it in a way that suits their communicative and social goals' (Galasiński and Kozłowska 2013, 280). As respondents tend to account for the *expectation* of the administrator (Norenzayan and Schwarz 1999), they put the researcher 'into the subject position of the addressee' (Galasiński and Kozłowska 2013, 3518).

This communicative dynamic was overt in both June's and Amy's narrative. As June formulated it, the questionnaires allowed the administrator to draw a 'profile' from the symptoms. Importantly, this communicative factor is not neutral: as June and Amy both feared the verdict of others, this tapped into their experienced symptoms beyond a mere problem of fit or accuracy of scoring. As the questionnaire administration impacted their fear (albeit with different outcomes), the questionnaire administration thus affected their *level* of experienced symptoms, rather than neutrally registering existing symptoms.

Thirdly, the case comparison showed that questionnaire administration can affect the *presence* of symptoms. As the second theme highlighted, the regular administration of questionnaires facilitated a process of self-evaluation. As patients are asked to regularly score a multitude of items on symptoms that they had not explicitly experienced before, the questionnaires prime awareness of symptoms that may have been less prominent or even absent. June's experience of suicidal thoughts is exemplary for the probing of experiences that were absent but were suggested by the (repeated) items in the questionnaires.

These findings are in line with the observation that *change research* can yield response shifts (Schwartz and Rapkin 2004; Westerman et al. 2008). This phenomenon occurs when an intervention, such as psychotherapeutic treatment, not only *reduces* initial symptoms but also changes the interpretation of symptom-focused items. Consequently, a pre-post change score cannot straightforwardly be interpreted as reduction of initial symptoms, but rather as *reprioritization, recalibration*, or *reconceptualization* of symptoms. Whereas response shift literature describes changed responses on measures as a consequence of an intervention, our findings suggest that measurement may function as an intervention *itself*. That is, interpretation of items may be facilitated by the *act* of scoring the items in measures themselves.

As June and Amy experienced questionnaires as a normative process of self-evaluation (theme 3), they were inclined to adjust their previous interpretation of their experienced symptoms (theme 4). Amy reprioritized feelings of guilt, which she previously did not interpret as symptom before. Nonetheless, upon being asked, she becomes aware of this feeling and thus scores the item higher than at pre-measurement, even though the *level* of experienced guilt did not change. An opposite change is that Amy always took her suicidal ideations as a sign of craziness, but her observation that they are on the 'low' side of response scales inclined Amy to recalibrate her perceived suicidality.

For June, the act of questionnaire administration had the opposite effect. The fact that her symptoms *could* fit so well confirmed her fears. Whereas incidental suggestions previously sparked anxious thoughts, the repetitive confrontation with the exhaustive list of possible symptoms facilitated manifest phobic thoughts. This way, she *recalibrated* (Schwartz and Rapkin 2004) her feared 'symptoms' from incidental and seemingly random into the 'fact' that she is psychiatrically ill. This further impacted the scanning of her own behavior and thoughts, which was unfocused and general at the start, but focused on concrete signals over time. Thus, she *reconceptualized* (Schwartz and Rapkin 2004) her fear of diseases as a manifest problem. For example, while she had not experienced suicidal ideations, being confronted with these items makes her aware of the fact that she *might* develop suicidal tendencies. The act of questionnaire administration thus affected her understanding of experienced symptoms such that it **worsened.*[5] the symptoms.

Our exploration of questionnaire administration *within* a clinical context adds an important feature to existing cognitive interview studies. Self-report symptom measurement is often advocated as a neutral tool to mirror existing symptoms. However, the fourth theme showed that symptoms rather started to mirror the questionnaire items. Particularly for June, it was the fact that her symptoms *might* fit with the questionnaires that the questionnaires turned into a tangible confirmation of her fears. For Amy, in contrast, the misfit with certain questionnaires and labels on the response scales, facilitated a more nuanced interpretation of her own symptoms. Moreover, the *absence* of assessed symptoms made her realize that other people could have worse symptoms. So, whereas for Amy, the impact of questionnaire administration changed her fears positively, for June, it had adverse effects. The questionnaires thus introduced a change in the understanding of the core symptoms that got them to seek treatment, in a way that would not (necessarily) have happened in treatment otherwise.

The findings from this qualitative study pose both an epistemic and an ethical problem for the use of questionnaire administration in psychotherapy research. Epistemically, in psychotherapy research, quantitative data are collected in order to infer *treatment efficacy*, operationalized as pre-post symptom reduction. However, as our findings indicate, at least part of pre-post changes may be due to the act of measurement. Even though there may be a treatment effect involved, it is not necessary nor sufficient to straightforwardly assume that the outcomes evidence treatment effect. Although part of the decrease of symptoms could be due to treatment, the overall outcome thus cannot validly be attributed to treatment per se.[6]

This epistemic problem substantiates the importance of adding a qualitative layer to psychotherapy research (Ponterotto 2005). Traditionally, qualitative methods (i.e., cognitive interviewing) are incorporated in the development of measures for the purpose of validation (Peterson, Peterson, and Gilmore Powell 2017). However, limiting qualitative research to the development phase of the measure may give users the impression that all possible problems with regard to concrete administration and interpretation were 'covered' during development and therefore do not require further attention (Moss 2013). By pursuing a qualitative analysis of the explicit process of questionnaire administration, the current study demonstrates that the meaning-making processes in questionnaire administration are neither completed nor fully covered. Rather, this study illustrates the contextual and idiosyncratic nature of questionnaire scores, especially when interpreted in the context of treatment effect.

As the studied process of questionnaire administration *did* facilitate clinical change, however, researchers may be inclined to infuse treatment with regular administration of questionnaires, rather than with qualitative assessment. In clinical practice, routine outcome measurement (ROM) is already pursued as a therapeutic tool (Boswell et al. 2013). Problematically, though, the impact of questionnaire administration on assessed symptoms is not observable *in* the numbers themselves. That is, the numbers do not show whether they actually capture the intended construct or are *shaped* by the act of measurement itself. Whereas for Amy, the therapeutic change and the impact of measurement were distinct processes which together facilitated clinical change, for June the measurement impact arguably came in the way of benefitting from therapeutic change (cf., Truijens, Desmet, De Coster, Uyttenhove, Deeren, Meganck, 2019). Therefore attention to measurement impact is not just vital for the epistemic validity of research, but it also affects the *ethics* of clinical research and practice. It is crucial to note that the impact of questionnaires is not *necessarily* beneficial and thus should be considered carefully in terms of the individual effects for individual patients.

Moreover, as the understanding of deterioration and adverse therapeutic change is very limited, researchers have called for less generalized and more concrete examples (De Smet et al. 2019). This requires the incorporation of a qualitative stance in psychotherapy research. As Hesse-Biber (2010) argues, ‘the method is but the tool; the methodology determines the way in which the tool will be utilized’ (Hesse-Biber 2010, 17). The advantage of taking a qualitative approach in exploring the implications of administering quantitative symptom measures is that it motivates researchers to stay attentive to how participants try to tell their story *via* questionnaires, rather than taking the face value of numerical data for granted.

This point further emphasizes the value of case comparison studies as part of psychotherapy research. Although their findings are not directly generalizable to a broader population, the case comparison does provide an in-depth reflection of clinical processes. The current study showed that despite initial similarities in cases, idiosyncrasy was present throughout the clinical processes with completely opposite outcomes. Comparison of multiple cases can ‘provide a relative standing to other similar cases’ and as such ‘connect [cases] to form a coherent body of knowledge’ (Iwakabe and Gazzola 2009, 604). The case comparison design can therefore be considered a *bottom-up synthesis* (Iwakabe and Gazzola 2009) of idiosyncratic clinical experiences, which allows for hypothesis-generating research (Stiles 2009) that is both relevant for epistemic growth of clinical knowledge as for ethical conduct in clinical practice.

A limitation of the current case comparison study is the use of a mixed method assessment in which participants were involved in an extensive data collection process. To limit Hawthorne-effects (i.e., bias imposed by participants’ awareness that they are being observed; see Truijens 2019), this study separated therapist and researcher strictly (i.e., the researcher did not know about treatment, the therapist did not know about research, and this was explicitly communicated to participants). To limit the possible impact of interview participation (see Galasiński and Kozłowska 2010, 2013), in this study we used both asked-upon and spontaneous narratives (note that June only provided spontaneous narrative and written notes on questionnaires).

The routine administration of questionnaires may have affected the reported experience of questionnaire administration, for example regarding the opportunity to self-monitor. This impact, however, raises an important consideration for the use of ROM in therapeutic processes. As ROM is increasingly applied – and often mandatory – in clinical practice, it is important to acknowledge the possible impact on clinical processes. Boswell et al. (2013) note that ‘we cannot forget that […] clients want the highest quality of care, but they also want relatively brief scales with a high degree of face validity.’ (p. 6). However, face validity can be rather deceptive in treatment and research, especially when patients do *not* benefit or experience adverse effects from ROM. According to Boswell et al. (2013), ROM ‘can provide glimpses into the unique world-view of the patient and the window through which the therapist and patient can look together.' (p. 7). To utilize such glimpses into individual meaning-making, it is vital that clinicians are *aware* of individual and possible adverse effects of (routine) measurement within a treatment process. A strength of the current case comparison study is that it provides clinicians with narrative clinical examples that turn the clinician's attention from abstract general principles to the narrative nature of clinical practice (Truijens 2019).

In line with the argument that case studies can help bridge the so-called 'research-practice gap' (Iwakabe and Gazzola 2009), this case comparison study urges quantitative researchers, qualitative researchers and clinicians to join efforts in developing a better understanding of how measurement works in clinical practice. Even *one* case of deterioration related to measurement is enough to recall general clinical deontology. Indeed, it is the ethical and epistemic responsibility of both researchers and practitioners to use measurement cautiously and to remain attentive to the price of monitoring in clinical practice.

Notes

1 'June' and 'Amy' are pseudonyms. The study designs and proceedings were approved by the Ethical Board of the Ghent University Hospital in Belgium (Registration numbers B670201318127 (SCS) and EC/2015/0085 (RCT)). All patients gave written informed consent to collect, analyse and publish their individual data throughout and after treatment. All identifying information concerning the patient has been changed to protect confidentiality. The data are denoted by an anonymous participant code and the patients are referred to by a pseudonym.

2 The present study has been presented by the first author to a diverse range of psychological and philosophical audiences (see Truijens 2019, 158).

3 Purposive selection based on familiarity with cases allows case synthesis that deepens understanding and enhances knowledge building (cf., Iwakabe and Gazzola 2009; McLeod and Elliott 2011; Stiles 1993).

4 The term 'symptoms' is strongly associated with the medical model discourse in which symptoms are understood as the observable signs of an underlying (causal) disease state. In contrast, we – the authors – are all research-practitioners who adhere to a personal recovery paradigm that focuses on *personal experience* and meaning-making. To avoid emphasis on a experts' interpretation of medical symptoms with a lack of agency for 'patients', we use the terms 'experienced symptoms', 'complaints' or 'presenting problems'.

5 See Truijens (2019) for post hoc calculated reliable change indices consistent with recovery/improvement (Amy) and deterioration (June).

6 Further consideration of the validity of inferences from these data and the sufficiency of methodological 'warrants' in the data analysis phase is taken as central question in Truijens, De Smet, Desmet, and Meganck (in press).

Notes on contributors

Femke L. Truijens holds a PhD in Clinical Psychology and a Master's degree in Philosophy of Science. She works as assistant professor clinical psychology at Erasmus University Rotterdam (The Netherlands), and as post doc guest researcher at Ghent University (Belgium). Her research is focused on validity of research methods and data collection processes in psychotherapy research. Furthermore, she works in her private psychotherapeutic practice for gifted adults.

Kimberly Van Nieuwenhove holds a PhD in clinical psychology and works as postdoc researcher and lecturer at the department of Psychoanalysis and Clinical Consulting at Ghent University, where studies the nature of interpersonal relationships in complex trauma and the processes of change in treatment. She is specialized in qualitative and mixed-methods case study research. Besides her academic work, she works as a psychotherapist in her private practice.

Melissa M. De Smet holds a PhD in clinical psychology and works as a postdoc researcher and lecturer at Ghent University (Belgium). Her research is yields the question of conceptualization and measurement of psychotherapy outcome from the patient's perspective. She currently works as an assistant lecturer and develops further research on treatment experiences in youth mental health care.

Mattias Desmet is professor in clinical psychology at the Faculty of Psychology and Educational Sciences at Ghent University, Belgium. He is the author of The Pursuit of Objectivity in Psychology and Lacan's Logic of Subjectivity: A Walk on the Graph of Desire. He works at his private practice as a Lacanian psychoanalyst.

Reitske Meganck is professor in clinical psychology at the Faculty of Psychology and Educational Sciences at Ghent University, Belgium. Her research is focused on psychotherapy processes and outcome, and she is one of the founders of the Single Case Archive, a database for published single case studies. She further works as a psychoanalyst in her private practice.

ORCID

Femke L. Truijens http://orcid.org/0000-0003-0790-7003
Kimberly Van Nieuwenhove http://orcid.org/0000-0002-3750-1314
Melissa M. De Smet http://orcid.org/0000-0002-8753-1757
Mattias Desmet http://orcid.org/0000-0002-7954-6078
Reitske Meganck http://orcid.org/0000-0002-9052-0100

References

Beck, A. T., R. A. Steer, and G. K. Brown. 1996. *Beck depression inventory*. (2nd edn). San Antonio: Harcourt Assessment.

Borsboom, D., G. J. Mellenbergh, and J. Van Heerden. 2004. The concept of validity. *Psychological Review* 111 (4):1061–71. doi:10.1037/0033-295X.111.4.1061.

Boswell, J. F., D. R. Kraus, S. D. Miller, and M. Lambert. 2013. Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research* 25:6–19. doi:10.1080/10503307.2013.817696.

Chambless, D. L., and T. H. Ollendick. 2001. Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology* 52 (1):685–716. doi:10.1146/annurev.psych.52.1.685.

Creswell, J. W., and D. L. Miller. 2000. Determining validity in qualitative inquiry. *Theory into Practice* 39 (3):124–30. doi:10.1207/s15430421tip3903_2.

De Smet, M., R. Meganck, K. Van Nieuwenhove, F. Truijens, and M. Desmet. 2019. No change? A grounded theory analysis of depressed patients' perspectives on non-improvement in psychotherapy. *Frontiers in Psychology* 10. doi:10.3389/fpsyg.2019.00588.

Derogatis, L. R. 1992. *SCL-90-R Administration, scoring and procedures manual.* (2nd edn) ed. Towson, MD: Clinical Psychometric Research Inc.

Elliott, R. (1999). *Client change interview protocol.* Retrieved May 2015, from http://experiential-researchers.org/instruments/elliott/changei.html

First, M. B., R. L. Spitzer, J. B. Miriam Gibbon, and W. Williams. 2002. *Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition (SCID-I/P).* New York: Biometrics Research, New York State Psychiatric Institute.

Flyvbjerg, B. 2006. Five misunderstandings about case-study research. *Qualitative Inquiry* 12 (2):219–45. doi:10.1177/1077800405284363.

Galasiński, D., and O. Kozłowska. 2010. Questionnaires and lived experience: Strategies of coping with the quantitative frame. *Qualitative Inquiry* 16 (4):271–84. doi:10.1177/1077800409354068.

Galasiński, D., and O. Kozłowska. 2013. Interacting with a questionnaire: Respondents' constructions of questionnaire completion. *Quality & Quantity* 47 (6):3509–20. doi:10.1007/s11135-012-9733-0.

Hesse-Biber, S. 2010. Qualitative approaches to mixed methods practice. *Qualitative Inquiry* 16 (6):455–68. doi:10.1177/1077800410364611.

Hill, C. E., H. Chui, and E. Baumann. 2013. Revisiting and reenvisioning the outcome problem in psychotherapy: An argument to include individualized and qualitative measurement. *Psychotherapy* 50 (1):68–76. doi:10.1037/a0030571.

Iwakabe, S., and N. Gazzola. 2009. From single-case studies to practice-based knowledge: Aggregating and synthesizing case studies. *Psychotherapy Research* 19 (4–5):601–11. doi:10.1080/10503300802688494.

Koeter, M. W. J., and J. Ormel. 1991. *General health questionnaire, Nederlandse bewerking: Handleiding*. Lisse: Swets, Test Services.

Luborsky, L. 1984. *Principles of psychoanalytic psychotherapy. A manual for supportive-expressive treatment*. New York: Basic Books.

McLeod, J. 2001. An administratively created reality: Some problems with the use of self-report questionnaire measures of adjustment in counselling/psychotherapy outcome research. *Counselling and Psychotherapy Research* 1 (3):215–26. doi:10.1080/14733140112331385100.

McLeod, J., and R. Elliott. 2011. Systematic case study research: A practice-oriented introduction to building an evidence base for counselling and psychotherapy. *Counselling and Psychotherapy Research* 11 (1):1–10. doi:10.1080/14733145.2011.548954.

Meganck, R., M. Desmet, C. Bockting, R. Inslegers, F. Truijens, M. De Smet, R. De Geest, et al. 2017. The Ghent Psychotherapy Study (GPS) on the differential efficacy of supportive-expressive and cognitive behavioral interventions in dependent and self-critical depressive patients: Study protocol for a randomized controlled trial. *BioMed Central* 18:126. doi:10.1186/s13063-017-1867-x.

Moss, P. A. 2013. Validity in action: Lessons from studies of data use. *Journal of Educational Assessment* 50:91–98.

Norenzayan, A., and N. Schwarz. 1999. Telling what they want to know: Participants tailor causal attributions to researchers‘ interests. *European Journal of Social Psychology* 29 (8):1011–20. doi:10.1002/(SICI)1099-0992(199912)29:8<1011::AID-EJSP974>3.0.CO;2-A.

Peterson, C., A. Peterson, and K. Gilmore Powell. 2017. Cognitive interviewing for item development: Validity evidence based on content and response processes. *Measurement and Evaluation in Counseling and Development* 50 (4):217–23. doi:10.1080/07481756.2017.1339564.

Ponterotto, J. G. 2005. Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology* 52 (2):126–36. doi:10.1037/0022-0167.52.2.126.

Schwartz, C., and B. Rapkin. 2004. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes* 2 (1):16. doi:10.1186/1477-7525-2-16.

Schwarz, N. 1999. Self-reports: How the questions shape the answers. *American Psychologist* 54 (2):93–105. doi:10.1037/0003-066X.54.2.93.

Stiles, W. B. 1993. Quality control in qualitative research. *Clinical Psychology Review* 13 (6):593–618. doi:10.1016/0272-7358(93)90048-Q.

Stiles, W. B. 2009. Logical operations in theory-building case studies. *Pragmatic Case Studies in Psychotherapy* 5 (3):9–22. doi:10.14713/pcsp.v5i3.973.

Truijens, F. L., M. Desmet, E. De Coster, H. Uyttenhove, B. Deeren, and R. Meganck. 2019. When quantitative measures become a qualitative storybook: A phenomenological case analysis of validity and performativity of questionnaire administration in psychotherapy research. *Journal of Qualitative Research in Psychology* 1–44. doi: 10.1080/14780887.2019.1579287

Truijens, F. L. (2019). The story of ‘the data’. On validity of data and performativity of research participation in psychotherapy research. *Doctoral dissertation.*

Truijens, F. L., M. De Smet, M. Desmet, and R. Meganck. in press. Validity of data as precondition for evidence. A methodological analysis of what is taken to count as evidence in psychotherapy research. *Philosophy, Psychiatry, and Psychology.*

Truijens, F. L., S. Cornelis, M. Desmet, M. De Smet, and R. Meganck. 2019a. Validity beyond measurement: Why psychometric validity is insufficient for valid psychotherapy research. *Frontiers in Psychology* 10. doi:10.3389/fpsyg.2019.00532.

Van Nieuwenhove, K., F. Truijens, R. Meganck, S. Cornelis, and M. Desmet. 2019. The nature and change of interpersonal patterns in complex trauma: An evidence-based case study. *Psychological Trauma: Theory, Research, Practice, and Policy*. doi:10.1037/tra0000438.

Wampold, B. E., and Z. E. Imel. 2015. *The great psychotherapy debate. The evidence for what makes psychotherapy work*. New York: Routledge.

Westen, D. (2002). *Clinical Diagnostic Interview*. Unpublished manual, Emory University.

Westerman, M. J., T. Hak, M. A. G. Sprangers, H. J. M. Groen, G. van der Wal, and A.-M. The. 2008. Listen to their answers! Response behaviour in the measurement of physical and role functioning. *Quality of Life Research* 17 (4):549–58. doi:10.1007/s11136-008-9333-6.

Willig, C. 2012. *Qualitative interpretation and analysis in psychology*. Berkshire: McGraw Hill, Open University Press.