



Development and validation of the German translation of the views on inpatient care (VOICE-DE) outcome measure to assess service users' perceptions of inpatient psychiatric care

F. Klein, J. Schleithoff, M. Schulz, K. Webel, H. Foullois, M. Loehr & K.-T. Kronmueller

To cite this article: F. Klein, J. Schleithoff, M. Schulz, K. Webel, H. Foullois, M. Loehr & K.-T. Kronmueller (2023) Development and validation of the German translation of the views on inpatient care (VOICE-DE) outcome measure to assess service users' perceptions of inpatient psychiatric care, *Journal of Mental Health*, 32:3, 560-566, DOI: [10.1080/09638237.2022.2118691](https://doi.org/10.1080/09638237.2022.2118691)

To link to this article: <https://doi.org/10.1080/09638237.2022.2118691>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 14 Sep 2022.



[Submit your article to this journal](#)



Article views: 1063



[View related articles](#)



[View Crossmark data](#)

ORIGINAL ARTICLE



Development and validation of the German translation of the views on inpatient care (VOICE-DE) outcome measure to assess service users' perceptions of inpatient psychiatric care

F. Klein^a , J. Schleithoff^a, M. Schulz^{a,b,c}, K. Webel^a, H. Foullois^d, M. Loehr^{a,b} and K.-T. Kronmueller^a

^aDepartment of Psychiatry and Psychotherapy, LWL-Klinikum Gütersloh, Gütersloh, Germany; ^bDiaconic University of Applied Sciences, Bielefeld, Germany; ^cInstitute of Health and Nursing sciences, Martin-Luther-University Halle Wittenberg, Halle, Germany; ^dDepartment of Hospitals and Healthcare, Landschaftsverband Westfalen-Lippe, Münster, Germany

ABSTRACT

Background: Service user satisfaction in inpatient psychiatric care is often measured with instruments that have been designed by professionals, without involving the perspective of service users. Views on Inpatient Care (VOICE), developed in England, is among the first service-user-generated outcome measures which included service users' perspectives in the process.

Aims: In this study, we aimed to validate a German version of VOICE.

Methods: The original questionnaire was translated into German and validated using data collected from 163 inpatients undergoing treatment in a psychiatric hospital. The instrument was tested for its psychometric properties, focusing on measurements of reliability and construct validity. Also, we assessed the impact of demographic variables. Finally, factorial analyses were carried out to compare the underlying factorial structure to the English version.

Results: The analyses revealed a high internal consistency ($\alpha = 0.90$). No significant impact of demographic variables was observed. Factorial analyses indicated a one-factor structure which accounted for 40.39% of variance.

Conclusions: Psychometrical evaluation of VOICE-DE indicated the questionnaire to be a suitable tool to assess service users' personal experience with treatment and satisfaction in German.

ARTICLE HISTORY

Received 8 November 2021

Revised 28 July 2022

Accepted 16 August 2022

Published online 29 August 2022

KEYWORDS

VOICE; psychometric evaluation; user perception

Introduction

There is a growing awareness in the mental health care field about the importance of incorporating service users' views on their needs for care, emphasizing their active role in the process of service planning, development, implementation and assessment. Mental health care encompasses a highly complex range of interacting services and diversified interventions. The specific type, intensity and duration of care should be tailored to the clinical condition and specific needs of each service user (McGrath & Tempier, 2003). However, the connection between intervention delivery and service users' improvement is often not straightforward (McGrath & Tempier, 2003).

This leads to the conclusion that there may be other aspects of the inpatient treatment, aside from disorder-specific treatment plans, which may improve or hamper the process of recovery. In this context, the World Health Organization developed the concept of responsiveness as an indicator of performance of health systems in terms of meeting the individual needs of service users in care (Bramesfeld et al., 2007; World Health Report, 2000). Following this idea, service users' views of their own health

are increasingly considered important for deliverance of high-quality health care (Bombard et al., 2018; Thornicroft & Tansella, 2005). In the past, numerous instruments have been developed aiming to gather service users' assessments of the treatment results achieved. These instruments can roughly be divided into two different kinds.

On the one hand, patient-reported outcome measures (PROMs) aim to collect measures of health status or health-related quality of life from the service user's perspective rather than from the clinician's perspective (Kendrick et al., 2016). On the other hand, patient-reported experience measures (PREMs) have started to surface in the field of service-user-reported measures during the last few years. PREMs are designed to assess service users' needs and experiences during treatment (Kingsley & Patel, 2017). This concept may very well add to the idea of a more responsive evaluation of medical health care. Service user satisfaction in mental health is a widely-used and valuable indicator for the quality of mental health care (Boyer et al., 2009; Institut für Qualitätssicherung und Transparenz im Gesundheitswesen [IQTIG], 2019). To learn about the satisfaction of service users is therefore imperative in order to

be effective and efficient in mental health services (Palha et al., 2017).

In a recent meta-analysis, Fernandes et al. (2020) identified 75 PREMs in the field of mental health care published in total. Of these 75 PREMs, 24 were designed for inpatient and residential settings. Evidence of patient involvement in the conception varied between instruments (Fernandes et al., 2020).

While PREMs seem to represent a valid approach to assess responsiveness of mental health care, the number of available instruments (especially for inpatient settings and non-English speaking countries) is too low. In addition to an increased need for instruments in this field, these instruments should be brief and psychometrically robust, allowing a clear measurement of inpatient care quality and responsiveness of inpatient mental health care (Evans et al., 2012).

Over the past several years, Germany has also become increasingly aware of the importance of a more inclusive mental health care system. As a sign of progress, there have been some innovative developments. One example is the funding and creation of Recovery Colleges (RC). The concept of RC has been developed in England and firstly introduced in 2009. It has shown to be a huge success, with (to date) over 80 RCs operating worldwide (Thériault et al., 2020). RCs focus on co-designed and co-facilitated educational courses on mental health and recovery (Bourne et al., 2018). While the majority of RCs are still located in the United Kingdom, Germany has just recently started to open the first German RCs (Zuaboni et al., 2020). Additionally, concepts which involve former service users as experts for recovery, e.g. Experienced Involvement ("EX-IN", Utschakowski, 2012) have become more common in Germany in recent years. Still, these promising advancements in inpatient mental health care cannot hide the fact that we currently do not have robust tools to get relevant information about the situation of inpatient mental health treatment in Germany (e.g. PREMs).

In the process of translating PREMs for the use in non-English speaking countries, there are aspects beside the plain translation into the respective languages which need to be addressed in thorough studies. For example, Dimitri et al. (2018) analyzed different predictors for the length of stay in psychiatric inpatient units in several different countries. In their study, the authors could show that some patient characteristics were associated with either a higher or lower length of stay, depending on the country the patient was treated in. The authors found these differences could not be explained by individual patient characteristics, but argued that differences in the respective national contexts and clinical practice may be responsible for their findings, thereby underlining the importance of national studies for valid insights about factors which influence the treatment duration of service users. While many European countries have, generally speaking, a shared understanding of mental health and share treatment approaches, they differ in the way mental health systems are organized. While some countries have a community-oriented approach (e.g. England), others follow a predominantly hospital-based approach (e.g. France;

Gutiérrez-Colosía et al., 2019). Germany to date also has what is essentially a hospital-based approach when it comes to intensive psychiatric care (Salize et al., 2007). The observable differences in amount and availability of psychiatric hospital beds may directly influence the kind of treatment which can be implemented. Presumably, countries with a predominantly hospital-based approach and with a comparatively large number of available beds could thus be able to provide or even focus on care for voluntarily treated patients. At the same time, countries with a community-oriented approach will probably predominantly treat the more severe cases in hospitals, possibly leading to a higher rate of compulsory admissions. There is reason to assume that these factors lead to differences in the suitability of a certain instrument in each country, contributing to the need for country-specific validation studies.

One interesting example of PREM assessing user satisfaction in in-house psychiatric treatment is the Views of Inpatient Care (VOICE) questionnaire (Evans et al., 2012), which has already been successfully translated into other languages in the past (e.g. Palha et al., 2017).

Because of the described lack of validated, brief PREMs for inpatient mental health care in Germany, the aim of the present study was to create a German version of VOICE, the VOICE-DE, using scientific approaches during the translation process as well as during the psychometrical validation.

Materials and methods

Procedure

Data were collected in the department of psychiatry of the Landschaftsverband Westfalen-Lippe (LWL) hospital in Gütersloh between November 2020 and December 2020, including 13 wards with inpatient psychiatric settings, of which one is a closed ward. The study was approved by the administration of the psychiatric institution.

In the course of hospitalization, as part of the information process about data privacy during treatment, every service user was briefed about ongoing regular voluntary assessments of treatment quality. For the present study, VOICE-DE was added to the standard questionnaire of the psychiatric institution, which was offered to every service user who was present during data collection. Because this was considered a routine quality assurance procedure, no ethics committee vote was obtained. The study was carried out in accordance with the declaration of Helsinki (World Medical Association, 2013). Thirty-four percent of all service users who were admitted during our study period took part in the survey.

Assessment tools

In order to fill the current gap, we developed a German version of the VOICE questionnaire (Evans et al., 2012). VOICE seemed to be a potentially interesting tool because of its participatory methodology in the generation process.

VOICE was also chosen because it has been psychometrically validated. Both the original VOICE (Evans et al., 2012) as well as the Portuguese translation (Palha et al., 2017) showed satisfactory psychometric properties such as high validity and reliability. Evans et al. (2012) reported satisfactory test characteristics, with high internal consistency ($\alpha = .92$) as well as test-retest reliability ($\rho = .88$, CI .81–.95). The correlation of $r = .82$ between VOICE and Service Satisfaction Scale-30 (Greenfield & Attkisson, 1989) displayed a high criterion validity. Besides the psychometric validation in the main publication (Evans et al., 2012), additional statistical evaluation has been carried out (Wykes et al., 2018), suggesting a two-factorial structure (subscales “security” and “care”). A German (unpublished) translation of the original VOICE was also done in the past (Bendix, 2010), though there is unfortunately a lack of information about the translation process and about statistical validity of the translated instrument. The VOICE is a self-report questionnaire for perceptions of clinical care and consists of 19 items with a six-point Likert scale ranging from “strongly disagree” (1) to “strongly agree” (6). In addition, there is a possibility to fill in free texts at the end of each topic (admission, care and treatment, medication, staff, therapy and activities, environment, diversity). The sum score ranges from 19 to 114, with higher scores expressing an overall more negative assessment of treatment. In the present study, in addition to the VOICE questionnaire, gender and age were asked for on a separate page.

Translation of VOICE-DE

The translation of VOICE-DE was conducted using standard scientific procedures, aiming mainly at two different aspects: firstly, we tried to make sure that the translated items represent the original meaning as closely as possible. Secondly, the translated items were reexamined in regards to their suitability for the German users, in order to guarantee a maximum of cultural adaptation. Next, we used our insights to create another translation of the original VOICE to German.

As a first step, the original items were translated by a German native speaker with profound knowledge of the English language. This version was compared with the already mentioned, unpublished and untested translation by Bendix (2010) and discussed in a multi-disciplinary team of experts in the field of mental health (1 health scientist, 1 academic nurse, 1 psychiatrist, 1 psychologist). As a result, small adjustments were made. Since the items 18 “I feel able to practice my religion whilst I’m in hospital.” and 19 “I think staff respect my ethnic background.” were deemed as less suitable for the average service user, the wordings were slightly adjusted to “I feel able to practice important elements of my religion whilst I’m in hospital.” and “I think staff respect my ethnic-cultural background.”. In the next step, an independent native English-speaking translator with advanced German language skills but no specific knowledge of psychiatry or VOICE back-translated the items to English. The back-translation was then compared to the

original VOICE by the English native speaker and the other team members. We discussed these results to find specific features which may have led to different wordings in the back-translation (e.g. use of “clinic” vs. “ward”). Finally, we used our insights to create our final German translation from the original VOICE. This final version was again extensively discussed on a single item level, regarding the headings for the different parts and also with respect to the questionnaire as a whole. The whole translation process was aligned with Til Wykes, one of the authors of the original VOICE.

The resulting questionnaire was considered to be easy to understand and to complete. In a next step, it was validated as a part of the already mentioned ongoing routine quality reviews of the psychiatric institution.

Data analysis

The data was analyzed using the IBM Statistical Package for the Social Science for Windows 25.0 (IBM Corp., 2017). Similar to Evans et al. (2012), sumscores were calculated by totaling all items. In case of missing data, a pro-rated score was calculated for participants who filled out at least 80% of all items. Cases with less than 80% were not included in the analyses. The pro-rated score was calculated as a product of the average of the items reported and the total number of items (as suggested by Wykes et al., 2018). Item number 6 (“The staff gives me medication instead of talking to me.”) was scored reversely. Cronbach’s α (Cronbach, 1951) was used as an indicator for internal reliability of VOICE-DE and item-total-correlations were calculated. Exploratory analyses were conducted to assess the impact of demographic variables in terms of gender, age or type of admission using one-way ANOVA.

Next, factorial analyses were used in order to identify the underlying structure of the translated version and to compare it with the reported two-factorial structure of the original VOICE (Wykes et al., 2018), using principal component analysis. As a first step, the factorability of the 19 VOICE-DE items was examined. With a Kaiser-Meyer-Olkin value of .88, above the recommended value of .6 (Möhring & Schlütz, 2013; Tabachnick et al., 2019), and a significant Bartlett’s test for sphericity ($p < .001$), the prerequisites were fulfilled. After examining the prerequisites for the factorial analysis, the eigenvalue method was used to get a first impression of possibly underlying factors. Because of the possibility of overestimating the number of factors (Moosbrugger & Kelava, 2020), factors were also visually identified using the scree plot method. The resulting factorial structure was finally validated by parallel analysis (Moosbrugger & Kelava, 2020).

Results

A total of 214 service users took part in this study

While all 19 items had missing data, the number of missing values (out of all 214 service users) ranged from 21 (item 1)

Table 1. Demographic characteristics of service users that fulfilled at least 80% ($n = 163$) and differences in mean VOICE-DE scores by demographic group.

	<i>n</i>	%	Mean sum score	SD	CI 95%	<i>p</i>
Gender						.113
Male	67	41.1	35.87	11.03	33.10–38.65	
Female	63	38.7	39.56	12.61	36.69–42.42	
Diverse	1	0.6	19.00			
Not disclosed	32	19.6	36.66	10.12	32.64–40.68	
Age						.233
Under 25	18	11	42.54	11.35	37.16–47.91	
25–34	35	21.5	38.51	10.06	34.66–42.36	
35–44	23	14.1	36.84	14.97	32.08–41.59	
45–54	38	23.3	35.85	11.63	32.15–39.55	
55–64	31	19.0	37.35	10.61	33.26–41.45	
65 and older	17	10.4	34.58	10.74	29.05–40.11	
Not disclosed	1	0.6	19.00			
Setting						.64
in-house	157	96.3	37.43	11.70	35.60–39.27	
day hospital	6	3.7	35.16	10.12	25.78–44.56	
Discharge period						.101
November	65	39.9	35.52	12.06	32.69–38.35	
December	98	60.1	38.56	11.22	36.26–40.87	
Admission						.753
Voluntary	150	92.0	37.25	11.61	35.37–39.14	
Involuntary	11	6.7	37.53	11.07	30.58–44.48	
Not disclosed	2	1.2	43.50	21.92	27.20–59.80	

Note. Day hospital, in this context, depicts service users who, as a last stage of their stationary treatment, continue the standard in-house treatment on the same ward, but already sleep at home.

to 77 (item 18). Notably, the items 16 “I feel staff respond well when the panic alarm goes off.” (54 missings), 19 “I think staff respect my ethnic background.” (68 missings) and 18 “I feel able to practice my religion whilst I’m in hospital.” (77 missings) were left void by more than 25% of all participants.

163 service users responded to at least 80% of the VOICE items and were considered in further psychometric calculation. Table 1 represents demographic information.

First, we examined the reliability of the VOICE questionnaire. Cronbach’s α was .90, indicating a high internal consistency of the VOICE items. All items presented high item-total correlations (ranging from .26 to .78; $M = .56$; $SD = .13$).

One-factorial ANOVAs were conducted in order to compare sum scores of different age ranges and gender. No significant differences were found regarding gender, $F(3,159) = 2.02$, $p = .113$, age ranges, $F(6,156) = 1.36$, $p = .23$, setting, $F(1,161) = 0.22$, $p = .64$, discharge period, $F(1,161) = 2.71$, $p = .10$, or type of admission, $F(2,160) = 0.28$, $p = .75$. For further details, please see Table 1.

Factorial analysis

Following the calculations of Wykes et al. (2018), we examined how many factors underlie the German translation of VOICE.

In order to identify and compute composite scores for the underlying factors of the VOICE-DE, principal component analysis was conducted. Using the eigenvalue method, four factors could be identified with an eigenvalue greater than one, explaining 61.02% of the variance. However, since

this method often overestimates the number of factors (Moosbrugger & Kelava, 2020), the visual analysis of the scree plot was included in the process, suggesting a one-factorial solution. This solution was also supported by the determination of the factors by parallel analysis (randomly generated eigenvalue from 2-factorial solution larger than the respective value of the empirical data: Moosbrugger & Kelava, 2020). Table 2 shows the factor loadings on a general factor, explaining a total of 40.39% of the variance. All items except item 6 “Staff give me medication instead of talking to me” load sufficiently on the principal factor ($> .40$).

Discussion

The aim of our study was to develop a German translation of the VOICE, a PREM-instrument which was originally developed involving service users’ views on acute psychiatric inpatient care.

Similar to results reported by Evans et al. (2012), the VOICE-DE was found to have high internal consistency. Interestingly, in contrast to findings regarding VOICE, which indicate a two-factorial-structure (Wykes et al., 2018), the present analyses could show a different factorial structure for VOICE-DE. The four factors, which were identified by principal component analysis, explained about 60% of the variance. Nevertheless, further statistical evaluation suggested a one-factor-solution to be more sound. This ultimately led to the decision to drop the idea of a four-factor-solution, in favor of a single-factorial-structure with high internal consistency. The identified factor (personal experience with treatment and satisfaction) showed a high internal consistency ($\alpha = 0.90$) and accounted for around 40% of the variance. In comparison to Wykes et al. (2018), whose factorial structure explained 95% of the variance in their data, our findings indicate our model to have a comparatively lower overall fit. At this point, it is difficult to clearly pin down a specific reason for this. On the one hand, factorial solutions may be grounded in the particular sample. It is unclear whether the reported structure would be found again in another sample. Here, more studies on VOICE as well as VOICE-DE could help to get a more reliable impression. On the other hand, as mentioned before, cultural differences and different health systems may influence the examined underlying constructs and thereby lead to distinct statistical findings. Palha et al. (2017) could have added helpful information for this matter but their sample was too small to conduct a factorial analysis for their Portuguese version of VOICE. Further translations should include factorial analyses in order to gain further insight into this matter.

All but item 6 (“Staff give me medication instead of talking to me.”) sufficiently loaded on the factor. This is in line with other publications regarding the English and Portuguese version of VOICE, which also reported this specific item to have weaker psychometric properties. Palha et al. (2017) found a lower item-total correlation for this

Table 2. Factorial analysis of VOICE-DE.

item no.	item	M (SD)	corrected item scale correlation	factor loading
9	Ich vertraue darauf, dass das Personal eine gute Arbeit macht. [<i>I trust the staff to do a good job.</i>]	1.79 (.82)	.78	.84
17	Das Personal reagiert angemessen, wenn ich mich in einer Krise an sie wende. [<i>I feel staff respond well when I tell them I'm in crisis.</i>]	1.86 (.88)	.73	.79
8	Wenn ich Gesprächsbedarf habe, ist das Personal für mich da. [<i>Staff are available to talk to when I need them.</i>]	1.74 (.85)	.72	.78
10	Ich habe den Eindruck, dass das Personal versteht, wie meine Krankheit mich beeinträchtigt. [<i>I feel that staff understand how my illness affects me.</i>]	2.03 (1.03)	.70	.78
11	Das Personal behandelt mich mit Respekt. [<i>I feel that staff treat me with respect.</i>]	1.69 (.87)	.68	.75
7	Das Personal interessiert sich für mich. [<i>Staff take an interest in me.</i>]	1.87 (.92)	.67	.74
16	Das Personal reagiert angemessen, wenn Alarm in Krisensituationen ausgelöst wird. [<i>I feel staff respond well when the panic alarm goes off.</i>]	1.78 (.76)	.65	.71
15	Auf der Station fühle ich mich sicher. [<i>I feel safe on the ward.</i>]	1.74 (.89)	.58	.65
13	Ich finde Einzelgespräche mit dem Personal hilfreich. [<i>I find one-to-one time with staff useful.</i>]	1.78 (.79)	.60	.65
12	Die Aktivitäten auf der Station entsprechen meinen Bedürfnissen. [<i>I think the activities on the ward meet my needs.</i>]	2.46 (1.07)	.55	.60
3	Stationsvisiten sind für mich hilfreich. [<i>Ward rounds are useful for me.</i>]	2.10 (1.04)	.54	.60
2	Ich kann bei meiner Behandlung und Versorgung mitbestimmen. [<i>I have a say in my care and treatment.</i>]	2.08 (.97)	.52	.56
5	Ich habe die Möglichkeit, meine Medikamente und Nebenwirkungen zu besprechen. [<i>I have the opportunity to discuss my medication and side effects.</i>]	1.89 (1.01)	.52	.55
19	Ich denke, das Personal respektiert meinen ethnisch-kulturellen Hintergrund. [<i>I think staff respect my ethnic-cultural background.</i>]	1.78 (.89)	.50	.54
1	Ich fühle mich willkommen, als ich auf diese Station kam. [<i>I was made to feel welcome when I arrived on this ward.</i>]	1.78 (.93)	.48	.54
18	Während des Krankenhausaufenthaltes kann ich wichtige Elemente meiner Religion ausüben. [<i>I feel able to practice important elements of my religion whilst I'm in hospital.</i>]	2.18 (1.10)	.45	.51
4	Ich habe den Eindruck, dass meine Medikamente mir helfen. [<i>I feel my medication helps me.</i>]	2.13 (1.06)	.41	.46
14	Während meines stationären Aufenthaltes ist es gut möglich, mit Familie und Freunden in Kontakt zu bleiben. [<i>I find it easy to keep in contact with family and friends when I'm on the ward.</i>]	2.30 (1.17)	.39	.45
6	Das Personal gibt mir eher Medikamente anstatt mit mir zu sprechen. (rec) [<i>Staff give me medication instead of talking to me.</i>]	2.42 (1.36)	.26	<.40

Note. Items sorted in descending order relating to factor loadings.

item and Wykes et al. (2018) reported this item not to load highly onto either of their factor scales.

Interestingly, the number of missing values was not equally distributed over all items. Two of the three items which showed more than 25% missing values dealt with aspects of diversity (regarding ethnic and religious needs), while one was related to staff's behavior in situations where the panic alarm goes off. These findings may be a result of country-specific distinctiveness. One possible explanation for this could be that there is, in comparison with an urban sample of English society (as found in Evans et al., 2012), comparatively less awareness of asking for specific ethnic or religious needs in the context of psychiatric hospital treatment, so subjects may simply not have known what exactly might be meant by these items. It also seems noteworthy that the respective items regarding aspects of diversity were further discussed and slightly adjusted during the translation process. This could underline the dimension "diversity" to be somewhat standing out for both the authors and the involved service users and could be a sign that the rewording did not work as expected. Another explanation may be that the observed population of service users itself is less diverse, either because of regional population characteristics or because of a lack of access for people with diverse or migrant backgrounds. The observed service area covers

urban as well as rural areas, possibly leading to a comparably high percentage of white German service users with either Christian or atheistic backgrounds who did not feel like items regarding diversity are relevant for them. Unfortunately, since we did not collect data on this exact matter, we cannot pin down the specific reason for the described findings. The high number of missing values for the item regarding situations with panic alarms, however, can be explained by hospital characteristics: the majority of wards have open doors and therefore mainly host voluntarily treated service users with high compliance for which the panic alarms are seldom needed.

Contrary to findings by Palha et al. (2017), who examined a Portuguese version of VOICE, no significant differences in scores were measured in regards to age of the participants. Palha et al. (2017) argued that, because of a high rate of long-term hospitalized older service users in the observed psychiatric facility, the more satisfied older service users possibly had more chances to socialize and thereby better fulfill their basic needs. However, the authors conducted their study in a facility for older women with both long-term and acute psychiatric care, making the generalizability of their results questionable (and thereby explaining different findings). Also, in contrast to Evans et al. (2012), our results do not indicate differences in service user

satisfaction depending on the type of admission. This result could be explained by the fact that our sample only included a small portion (6.7%) of compulsorily treated service users. Even though this reflects the typical reality of the hospital, it is unclear whether a more balanced sample in this regard would have led to significant differences.

Limitations

Several limitations must be acknowledged. First of all, similar to Palha et al. (2017), but in contrast to Evans et al. (2012), data was collected at a single psychiatric facility, possibly limiting the generalizability of the findings. Therefore, future studies should aim to further validate VOICE-DE using a multi-institutional sample. Regarding the evaluation of psychometric properties, the assessment of test-retest reliability and the validation of our proposed one-factorial structure would still be of interest and should be considered in further studies. Another limitation is related to the participation rate of 34% in the present study. In addition to this, 25% of the service users who participated had to be excluded due to excessive missing items. Since data was collected as part of the routine quality reviews, these rates could just represent a generally low motivation to take part in (or to complete) evaluations. Also, the participation rate reported by Evans et al. (2012), 45%, seems to be somewhat comparable to the present study. At the same time, specific factors could have led to some kind of selection bias or response bias (e.g. because of a less severe impairment to health at the end of treatment or because of a particularly bad or good treatment experience). Additional data (e.g. type of diagnosis, educational/professional status, prior psychiatric treatments or length of stay) would have been helpful to gain insights into possibly influencing factors, both concerning the participation rate and the ratings itself. Therefore, future studies should reconsider these aspects. Regarding the sampling procedure, the present study slightly differed from Evans et al. (2012) and Palha et al. (2017). In contrast to both studies, we did not specify any inclusion criteria regarding the amount of time the service user had to be present on the ward. While the majority of voluntarily treated service users in the hospital of the present study are treated for more than 7 days, we did not assess the individual length of stay of each participant.

Finally, we did not include users' perspectives during the course of translating the original questionnaire. While the original questionnaire had been developed involving the feedback and advice from service users, the translation process may have altered the service user orientation to some extent. Therefore, further studies should include some sort of service user feedback measurements (e.g. additional items measuring feasibility and acceptability of the questionnaire).

Conclusion

To sum up, the present study provides validation data on the German translation of a PREM, the VOICE-DE questionnaire. The current work is, to our knowledge, among

the first publications ever to deliver a psychometrically evaluated German translation of a PREM whose original version has been developed collaboratively with service users and professionals. Our study answers the need for psychometrically tested, user-generated instruments for the assessment of quality and responsiveness of inpatient psychiatric mental health care in Germany. Further research could try to further develop the instrument. This could be achieved by further evaluating VOICE-DE, especially with a larger multi-institutional sample. While the current work is a first step, future research in this direction has the potential to be very rewarding.

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki. Since the data of the study was collected as part of the routine quality management process, no ethics approval was obtained. During their stay, patients are informed about the voluntary feedback questionnaires. The obtained data is completely anonymized. The local laws regulate that studies of this kind do not need specific ethics approval or written informed consent. For further details, please refer to <https://www.medizin.uni-muenster.de/ek/ethik-kommission/antragsunterlagen/unterlagen-fuer-antraege/ausnahmen.html>.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

F. Klein  <http://orcid.org/0000-0002-6449-2345>

Data availability statement

The data that support the findings of this study are not publicly available but are available from the corresponding author upon reasonable request.

References

- Bendix, S. (2010). *VOICE – Views on inpatient care – German unpublished version.*
- Bombard, Y., Baker, G. R., Orlando, E., Fancott, C., Bhatia, P., Casalino, S., Onate, K., Denis, J. L., & Pomey, M.-P. (2018). Engaging patients to improve quality of care: A systematic review. *Implementation Science*, 13(1), 98. <https://doi.org/10.1186/s13012-018-0784-z>
- Bourne, P., Meddings, S., & Whittington, A. (2018). An evaluation of service use outcomes in a Recovery College. *Journal of Mental Health*, 27(4), 359–366. <https://doi.org/10.1080/09638237.2017.1417557>

- Boyer, L., Baumstarck-Barrau, K., Cano, N., Zendjidian, X., Belzeaux, R., Limousin, S., Magalon, D., Samuelian, J. C., Lancon, C., & Auquier, P. (2009). Assessment of psychiatric inpatient satisfaction: A systematic review of self-reported instruments. *European Psychiatry*, 24(8), 540–549. <https://doi.org/10.1016/j.eurpsy.2009.05.011>
- Bramesfeld, A., Wedegärtner, F., Elgeti, H., & Bisson, S. (2007). How does mental health care perform in respect to service users' expectations? Evaluating inpatient and outpatient care in Germany with the WHO responsiveness concept. *BMC Health Services Research*, 7, 99. <https://doi.org/10.1186/1472-6963-7-99>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Dimitri, G., Giacco, D., Bauer, M., Bird, V. J., Greenberg, L., Lasalvia, A., Lorant, V., Moskalewicz, J., Nicaise, P., Pfennig, A., Ruggeri, M., Welbel, M., & Priebe, S. (2018). Predictors of length of stay in psychiatric inpatient units: Does their effect vary across countries? *European Psychiatry*, 48(1), 6–12. <https://doi.org/10.1016/j.eurpsy.2017.11.001>
- Evans, J., Rose, D., Flach, C., Csipke, E., Glossop, H., McCrone, P., Craig, T., & Wykes, T. (2012). Voice: Developing a new measure of service users' perceptions of inpatient care, using a participatory methodology. *Journal of Mental Health*, 21(1), 57–71. <https://doi.org/10.3109/09638237.2011.629240>
- Fernandes, S., Fond, G., Zendjidian, X. Y., Baumstarck, K., Lançon, C., Berna, F., Schurhoff, F., Aouizerate, B., Henry, C., Etain, B., Samalin, L., Leboyer, M., Llorca, P. M., Coldefy, M., Auquier, P., & Boyer, L. (2020). Measuring the Patient Experience of Mental Health Care: A Systematic and Critical Review of Patient-Reported Experience Measures. *Patient Preference and Adherence*, 14, 2147–2161. <https://doi.org/10.2147/PPA.S255264>
- Greenfield, T. K., & Attkisson, C. C. (1989). Steps toward a multifactorial satisfaction scale for primary care and mental health services. *Evaluation and Program Planning*, 12(3), 271–278. [https://doi.org/10.1016/0149-7189\(89\)90038-4](https://doi.org/10.1016/0149-7189(89)90038-4)
- Gutiérrez-Colosía, M. R., Salvador-Carulla, L., Salinas-Pérez, J. A., García-Alonso, C. R., Cid, J., Salazzari, D., Montagni, I., Tedeschi, F., Cetrano, G., Chevreul, K., Kalseth, J., Hagmair, G., Straßmayr, C., Park, A. L., Sfectu, R., Ala-Nikkola, T., González-Caballero, J. L., Rabbi, L., Kalseth, B., & Amadeo, F. (2019). Standard comparison of local mental health care systems in eight European countries. *Epidemiology and Psychiatric Sciences*, 28(2), 210–223. <https://doi.org/10.1017/S2045796017000415>
- IBM Corp. (2017). *IBM SPSS Statistics for Windows (Version 25.0) [Computer software]*. IBM Corp.
- Institut für Qualitätssicherung und Transparenz im Gesundheitswesen. (2019). *Methodische Grundlagen V1.1*. https://iqtig.org/dateien/dasiq-tig/grundlagen/IQTIG_Methodische-Grundlagen-V1.1_barrierefrei_2019-04-15.pdf
- Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., Bhattacharya, A., Gimson, A., Brütt, A. L., Jong, K., & de Moore, M. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *The Cochrane Database of Systematic Reviews*, 7, CD011119. <https://doi.org/10.1002/14651858.CD011119.pub2>
- Kingsley, C., & Patel, S. (2017). Patient-reported outcome measures and patient-reported experience measures. *BJA Education*, 17(4), 137–144. <https://doi.org/10.1093/bjaed/mkw060>
- McGrath, B. M., & Tempier, R. P. (2003). Implementing quality management in psychiatry: From theory to practice – Shifting focus from process to outcome. *The Canadian Journal of Psychiatry*, 48(7), 467–474. <https://doi.org/10.1177/070674370304800707>
- Möhring, W., & Schlütz, D. (2013). *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Springer. <https://doi.org/10.1007/978-3-531-18776-1>
- Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion*. Springer. <https://doi.org/10.1007/978-3-662-61532-4>
- Palha, J., Palha, F., Dias, P., & Gonçalves-Pereira, M. (2017). Portuguese adaptation and input for the validation of the views on inpatient care (VOICE) outcome measure to assess service users' perceptions of inpatient psychiatric care. *Acta medica Portuguesa*, 30(11), 790–795. <https://doi.org/10.20344/amp.8596>
- Salize, H. J., Rössler, W., & Becker, T. (2007). Mental health care in Germany: Current state and trends. *European Archives of Psychiatry and Clinical Neuroscience*, 257(2), 92–103. <https://doi.org/10.1007/s00406-006-0696-9>
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Thériault, J., Lord, M.-M., Briand, C., Piat, M., & Meddings, S. (2020). Recovery colleges after a decade of research: A literature review. *Psychiatric Services*, 71(9), 928–940. <https://doi.org/10.1176/appi.ps.201900352>
- Thornicroft, G., & Tansella, M. (2005). Growing recognition of the importance of service user involvement in mental health service planning and evaluation. *Epidemiologia e psichiatria sociale*, 14(1), 1–3. <https://doi.org/10.1017/s1121189x00001858>
- Utschakowski, J. (2012). EX-IN-Ausbildungen: Experienced involvement-Pro & Kontra. *Psychiatrische Praxis*, 39(5), 202–203.
- World Health Report. (2000). *Health systems: Improving performance*. World Health Organisation.
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>
- Wykes, T., Csipke, E., Rose, D., Craig, T., McCrone, P., Williams, P., Koeser, L., & Nash, S. (2018). Patient involvement in improving the evidence base on mental health inpatient care: The PERCEIVE programme. *Psychological Medicine*, 48(3), 488–497. <https://doi.org/10.3310/pgfar06070>
- Zuaboni, G., Schulz, M., & Löhr, M. (2020). Recovery Colleges: Gemeinsam psychische Gesundheit und soziale Teilhabe durch Bildung fördern. *Kerbe*, 1, 25–27.