J. Psychosoc. Rehabil. Ment. Health (2021) 8:159–169
https://doi.org/10.1007/s40737-021-00216-w

ORIGINAL ARTICLE

Mini-ICF-APP Inter-Rater Reliability and Development of Capacity Disorders Over the Course of a Vocational Training Program—A Longitudinal Study

M. Burri · L. P. Werk · A. Berchtold · M. Pugliese · B. Muschalla

Received: 5 January 2021 / Accepted: 5 April 2021 / Published online: 22 April 2021

Abstract The Mini-ICF-APP is an established instrument in social medicine, especially in the context of work ability assessment. In this study, the 13 capacity dimensions of Mini-ICF-APP were tested for its inter-rater reliability in the context of a vocational training program for persons with chronic and mental health problems in Switzerland. Also, the development of capacity impairments was investigated over the course of the long-term vocational training programs.

61 training reports on chronically ill persons with mental health impairments were collected within a vocational training program in 2018–2019. Capacity impairment of the trainees were assessed at the beginning of the intervention (t0), after three months (t1) and after 6 to 9 months (t2) by a job attendant and a consultant at each of the three time points. Inter-rater reliabilities for each time point have been calculated. Development of capacity impairment over the course of the vocational training were investigated by variance analysis with repeated measurements.

The inter-rater reliability increased in all 13 Mini-ICF-APP capacity dimensions over the course from t0 to t2. Spearman correlation in each capacity dimension reached sufficient values (r = 0.55*–0.97**). There was no statistically significant change of capacity impairment over the course of the vocational training. Ten capacity dimensions showed a decreasing tendency, three showed a tendency to improve capacity levels.

Through repeated application and training programs the capacity raters seem to be able to improve reliability of their assessments. The phenomenon of slight increase of capacity impairments over the time may be due to the fact that context and rater's knowledge on participants and context demands changed over the time.

Keywords Capacities · Soft skills · ICF · Mini-ICF-APP · Professional training

M. Burri and L. P. Werk shared first authorship.

M. Burri · M. Pugliese
Federal Disability Insurance Office of the Canton of Fribourg, Fribourg, Switzerland

L. P. Werk (✉) · B. Muschalla
Institute of Psychology, Psychotherapy and Diagnostics, Technische Universität Braunschweig, Humboldtstr. 33, 38106 Braunschweig, Germany
e-mail: l.werk@tu-bs.de

A. Berchtold
University of Lausanne & NCCR LIVES, Lausanne, Switzerland

Introduction

Inter-rater Reliability In Social Medicine and Work Ability Assessment

The assessment of work ability in social medicine requires observation and exploration data [1]. Work ability cannot simply be judged by asking the patient. But especially observation data vary depending on the individual decisions within an interviewer's assessment, or by the assessment setting (outpatient assessment of one hour versus inpatient assessment with observation data over 5 weeks from different professions). To minimize variance in observation-based data interpretation and to get more objective standards for work ability measurement, inter-rater reliability is needed. The inter-rater reliability compares the assessment of two or more raters in the same clinical cases [2].

Inter-rater reliability has been a part of the evaluation of several participation-oriented instruments in recent years. For example: in the measurement of Activities of Daily Living (ADL; [3] and Global Assessment (CGAS; [4]) for children and in elderly care a fair to high inter-rater agreement of $r = 0.27$ to $r = 0.94$ were reported [5–7]. Lenze et al. [8] found high intra-class correlations for The Pittsburgh Rehabilitation Participation Scale (PRPS).

Inter-rater reliabilities are routinely assessed for symptom observation. For example, reliabilities of instruments assessing obsessive–compulsive symptoms (AMPD Rating Scale; [9] and disorders (Y-BOCS; [10], show inter-rater reliabilities of $r = 0.47$ to $r = 0.93$ [9, 11]. For the Children's Depression Inventory [12] and the German structured diagnostic interview for mental disorders in children and adolescents (Kinder-DIPS; [13]) high kappa statistics between $r = 0.86$ and $r = 0.90$ were reported [14].

In assessment of social performance and work ability there are also some measures with reported inter-rater reliabilities. Schaub, Brüne, Jaspen, Pajonk, Bierhoff and Juckel [15] investigated the inter-rater agreement in the Personal and Social Performance Scale (PSP; [16]) in a sample of chronic schizophrenia patients and found ICC between $r = 0.54$ and $r = 0.82$ [15]. In a validation study of MELBA [17], a measurement tool for vocational rehabilitation and inclusion, Achterberg, Wind, Prinzie and Frings-Dresen [18] reported poor to moderate and some excellent ICC. For Assessment of Life Habits (LIFE-H; [19]) for older adults Noreau et al. [20] found poor and good ICC between $r = 0.30$ and $r = 0.97$ [20].

Reviewers have criticized the low inter-rater reliabilities in several studies within the social medicine practice [21]. Low agreements in inter-rater reliability could arise from different or undefined reference contexts serving as a rule [1]. The question of how naturalistic the conditions for the assessment of work ability are e.g. in a clinical setting has to be kept in mind. Nevertheless, the examination of objectivity constitutes an important aspect of test accuracy and compliance with the quality standards of psychometric tests [22].

The Mini-ICF-APP – an Instrument for Objective Work Ability Assessment

The Mini-ICF-APP is one of the leading measurement tools for work ability assessment and permits a systematic description at the level of disabilities and impairments [23]. It was adapted from the structure of the ICF and combines functions, capacities and participation in an interactive multidimensional construct [24]. The Mini-ICF-APP was validated several times in German [25], Italian [26], English [27] and Polish language [28]. Psychological capacities and participation (impairment) can be described by using a semi-structured interview on 13 capacity dimensions: *adherence to regulations*, *planning and structuring of tasks*, *flexibility*, *applying expertise*, *capacity to judge and decide*, *endurance*, *assertiveness*, *contacts with others*, *teamwork capacity*, *self-care*, *mobility*, *proactivity* and *familiar and intimate relationships* [25, 27]. Usually the Mini-ICF-APP is used for measuring work ability in the context of an existing present workplace or potential workplaces on the general labour market depending [1, 25, 27]. The Mini-ICF-APP and its capacity dimensions are now the core contents in the AWMF guidelines for social medicine assessment of mental disorders [29]. Additionally, a self-rating version and a version for assessment of capacity demands of workplaces has been developed [30, 31].

Interrater-Reliability of the Mini-ICF-APP

In the Mini-ICF-APP validation studies inter-rater reliabilities were reported as well. Some critical voices described the Mini-ICF-APP as "unsuitable

instrument" [32] with low inter-rater reliabilities [33], like $r = 0.43$ in the study of Kunz et al. [34]. But, to fix the problem of divergent ratings due to different contexts of references, Linden et al. [25] had conceptualized a training for Mini-ICF-APP assessment, which increased the inter-rater reliability of $r = 0.70$ up to $r = 0.92$ [25]. Manual and exploration guidelines give the basis for Mini-ICF-APP ratings [25, 31]. The important basis for reaching good inter-rater reliability is that raters refer to the same and clearly defined context. The rating can only be as good as the reference contexts are defined and the raters refer to these definitions when making their ratings. The reference context must be defined before the exploration and the rating. Contexts may be for example a specific workplace with specific capacity demands, or the general labor market which requires a basic level in all capacities, or living on one's own with the demands of daily duties of housework and general life activities [31].

There are several studies, which reported good to excellent inter-rater reliabilities of the Mini-ICF-APP. Muschalla [30] found high kappa between $r = 0.71$ (*endurance*) and $r = 0.94$ (*mobility*) for the German version. Balestrieri et al. [26] revealed ICC between $r = 0.79$ and $r = 0.98$ in the Italian validation study. The weakest ICC was found for *mobility* [26]. Molodynski et al. [27] reported a mean score of $r = 0.89$ for the English version of the Mini-ICF-APP. The authors of the Polish Mini-ICF-APP study found ICC between $r = 0.59$ (*resistance* and *endurance*) and $r = 0.80$ (*competence to judge and decide*, *proactivity* and *spontaneity*) [28]. Inter-rater reliability can increase the longer the raters are trained [25].

Training Effects in Professionals

The principle of "learning by doing" and gaining experience over time is applied in many occupational fields [35]. Psychotherapists and physicians carry out practical training [36, 37]. Meta-analyses by Dush, Hirt and Schroeder [38] and Lyons and Woods [39] found significant effects between the practical work experience of psychotherapists and the effectiveness of the treatment of child behavior disorders [38] and rational-emotive therapy [39]. Dauwerse, Stolper, Molewijk and Widdershoven [40] emphasize the importance of practice over time in the training of health care professionals.

Training effects over time were also found in the assessment of psychological issues. Warshaw, Dyck, Allsworth, Stout and Keller [41] conducted a long-term study to test the inter-rater reliability with measurement points after 1, 6 and 12 months with the Longitudinal Interval Follow-up Evaluation (LIFE). The authors compared new raters with experienced raters and found tendentially higher intra-class correlation coefficients for the experienced raters at all measurement points [41]. With increasing experience of raters, the focus on situation-specific features, the contextual reference as well as the interpretation of observed behavior in contrast to behavior descriptions increases [42, 43].

Vocational Training

Until now, several research studies have shown the usefulness of vocational trainings: Lysaker, Davis, Bryson and Bell [44] examined the effect of a vocational rehabilitation program (Indianapolis Vocational Intervention Program (IVIP)) for patients with schizophrenia spectrum disorders compared with the usual service in job placement. Participants in the rehabilitation program found a job significantly faster and generally performed better in the workplace than participants of the control group. An RCT study was also conducted for the application of vocational rehabilitation programs for affective disorders. Bejerholm, Larsson and Johanson [45] reported a higher effectiveness of a disorder-specific vocational rehabilitation program (in this case Individual Enabling and Support (IES)) compared to traditional vocational rehabilitation. In addition to the higher rate of employment and higher number of working hours per week, the depression scores were significantly reduced. A controlled study by Watzke, Galvao and Brieger [46] showed a significantly higher employment rate, a reduction in symptoms and a subjectively higher level of well-being and functionality in the 9-month follow-up after a vocational rehabilitation program in patients with various mental disorders compared to the control group. In two controlled, randomized studies with patients of different mental disorder groups, Wallace and Tauber [47] emphasized the effectiveness of workplace-related skills training. With the vocational skills trainings, patients were more likely to find a workplace, worked a higher number of hours and earned more than patients

without workplace skills training [47]. An RCT study by Berglund et al. [48] compared the employment rate between the two vocational rehabilitation programs Multidisciplinary assessments and individual rehabilitation management and Acceptance and commitment therapy (ACT) and a control group. Participants in the training programs were able to return to their jobs significantly more often and reported having increased employability compared to the control group [48]. A systematic review by Michon et al. [49] reports the person-related predictors of employment outcomes after participation in vocational rehabilitation programs. In 8 of 16 studies, better work performance, higher self-efficacy and increased social functionality were identified as strengthened factors after a vocational rehabilitation program [49]. In the field of early psychosis, a systematic review of the effectiveness of early intervention programs for employment was generated by Bond et al. [50]. In the eleven studies, 29% of patients were employed with the usual support, the employment rate among vocational services patients was 49% [50]. In the rehabilitation of patients after acquired brain injury, a review of 12 studies by Donker-Cools, Daams, Wind and Frings-Dresen [51] showed that workplace training, skills training and education are effective vocational rehabilitation programs.

In Switzerland, persons with chronic mental health problems and problems with work ability can participate in longitudinal vocation reintegration programs. The here investigated program is individualized according to the concrete health and participation problems of the person. The participants of the program are integrated in companies and work in concrete workplaces which fit their capacity level. Participants are monitored and supported by social work professionals and physicians over the course. Participant's capacity level (and impairments) are monitored over the course of the longitudinal program (in this present investigation: from 2018–2019).

Question of Research

This study was the first evaluation of the Mini-ICF-APP in Switzerland and included three points of measurement within vocational trainings. For a reliable insurance-medicine assessment in questions of invalidity and work ability, the Mini-ICF-APP is a fixed component in Switzerland since 2014. As a part of Switzerland's social security law [52], it is indispensable to examine the fulfilment of quality standards and correspond assessments of different reviewers in the Mini-ICF-APP. This present study took on this task and investigated quantitative and qualitative changes in the assessment of the Mini-ICF-APP ratings across three points of measurement. Instead of using the mean score, the inter-rater reliability was calculated per each of the 13 capacity dimensions itself. The capacity assessment was done with participants of a vocational training program which was several months of duration.

The questions of research were therefore:

1. What is the inter-rater reliability of the 13 capacity dimensions of the Mini-ICF-APP in a Switzerland naturalistic vocational training setting?
2. (How) Do the inter-rater reliabilities of the capacity ratings change across three measurement points?

Methods

A sample of training reports on 61 vocational training participants were investigated. They have been collected from three different disability insurance institutions in Switzerland (CIS, Ritec, Sonora) between January 2018 and August 2019. An assessment of the capacity dimensions of the Mini-ICF-APP was carried out at the beginning of the intervention (t0), after three months (t1) and after six to nine months (t2). The vocational training was done with patients with chronic illnesses of diverse types, e.g. musculoskeletal diseases, malignant tumors, diabetes, organ damage, with resulting mental impairments and mental disorders. The 13 capacity dimensions of the Mini-ICF-APP were rated by two different raters in each training report. The raters had two different professional backgrounds: The Swiss Disability Insurance employs consultants and job attendants. Consultants are integration specialists who inform and advise insured persons about disability insurance benefits and support them in their vocational integration process in the cantonal office. The job attendants are Master Social Professionals who supervise the vocational integration process in the training centre. Within the vocational training, both professionals are focusing on capacity training. On behalf of the disability insurance,

integration programs for vocational rehabilitation are carried out in the training centers. The raters conduct these training programs and are available to advise the patients during the intervention. The training contents simulate the specific field of work and aim for the (re)development of job-relevant capacities. Therapeutic methods, e.g. learning relaxation techniques, are also used.

Each rater completed a rater training for Mini-ICF-APP inspired by the manual [25]. The training included the meaning of accurate description of the reference context, the assessment of capacities and rules of observation. The raters assessed the capacity impairments in respect to the reference context "work on the general labour market" in a semi-structured interview for each of the 13 capacity dimensions [25]. In the sense of a naturalistic study, the raters based their assessment of the capacity impairments on general employability. Each of the 13 capacity dimensions was rated from 0 = *no impairment*, 1 = *mild impairment*, i.e., there are some difficulties for the person to fulfill the demands, but there are no negative consequences, 2 = *relevant impairment*, i.e., there are visible problems in fulfilling the demands, 3 = *severe impairment*, i.e., help from others is needed regularly in order to fulfill the demands and activities to 4 = *full impairment*, i.e., no respective activity is possible and complete dispensation is necessary. Sources of information were the observation during the interview and the information of capacity impairment explored from the patient. Exploration was done behavior-oriented.

Spearman correlations for the inter-rater reliabilities and an analysis of variance with repeated measurement for the analysis of significant changes in the assessment of capacities over the vocational training program were calculated. As there was no confirmatory research question for this naturalistic sample, no sample size was calculated in advance. According to McHugh [2], correlative measures for determining inter-rater reliabilities such as kappa or Spearman correlations can already be used with a small sample size of 5 participants or more. Sim and Wright [53] found that between two raters in a dichotomous variable, a significant kappa value with a power of 80% can be achieved in a one-tailed test when there are between 8 and 39 participants, depending on the level of the kappa value.

Since not all participants could be assessed at all three points of measurement, missing values were replaced by mean values when calculating the capacity ratings over the course of the vocational training program. The mean value imputation was performed by using the mean value from the ratings of the same participant at a different point in time (mean value of the time series) for calculation. To calculate the capacity ratings development over the course of a vocational training program, the mean value of both ratings was used for each point of measurement (Figs. 1, 2, 3 and 4).

Results

At the first measurement point (t0) two raters were available in 19 cases. At t1 there were 17 and at t2 14 cases which included ratings from two raters. The capacity impairments were rated by consultants and job attendants (Table 1).

Inter-rater reliabilities were low at the first time point of measurement and increased over the course from t0 to t2 (Table 2). All 13 dimensions showed high significant correlations at t2 (r = 0.55*–0.97**). The highest concordance was found for *adherence to regulations*, *mobility* and *proactivity* (> 0.90**).

None of the 13 capacity dimensions increased or decreased statistically significant in capacity impairment over the course from t0 to t2. In ten out of 13 capacities (*adherence to regulations*, *endurance*, *applying expertise*, *capacity to judge and decide*, *assertiveness*, *contacts with others*, *teamwork*

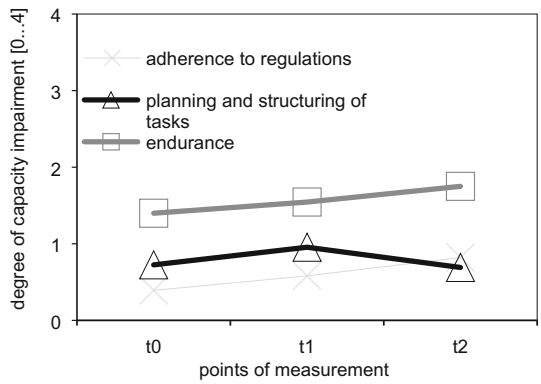

Fig. 1 Development of capacity impairment in the dimensions *adherence to regulations*, *planning and structuring of tasks* and *endurance* over the course of the vocational training (N = 61)

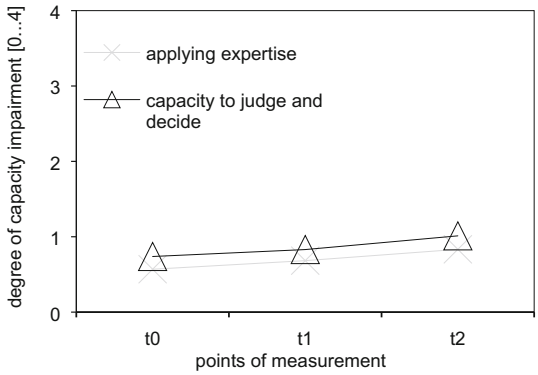

Fig. 2 Development of capacity impairment in the dimensions *applying expertise* and *capacity to judge and decide* over the course of the vocational training ($N = 61$)

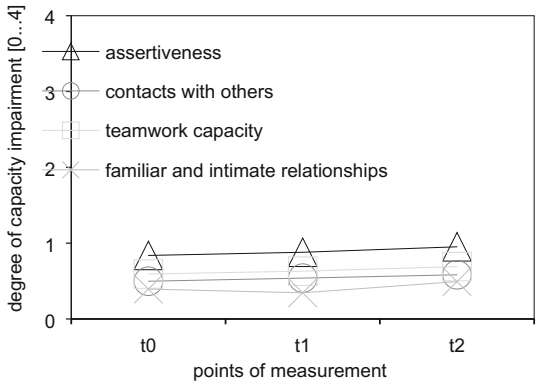

Fig. 3 Development of capacity impairment in the dimensions *assertiveness*, *contacts with others*, *teamwork capacity* and *familiar and intimate relationships* over the course of the vocational training ($N = 61$)

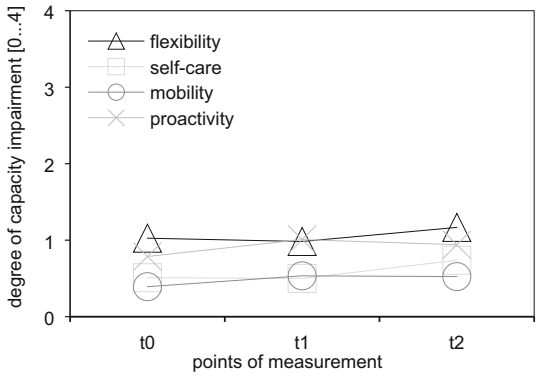

Fig. 4 Development of capacity impairment in the dimensions *flexibility*, *mobility*, *self-care* and *proactivity* over the course of the vocational training ($N = 61$)

capacity, *familiar and intimate relationships*, *self-care*, *proactivity*) there was a tendency of increasing impairment. Only three capacities (*planning and structuring of tasks*, *flexibility*, *mobility*) had tendentially decrease in impairments (Figs. 1, 2, 3 and 4). The high number of capacities with increasing impairment (instead of declining) is contradictory to the idea that the occupational training should improve the patient's capacity levels.

Discussion

First, results of the study show that the inter-rater reliability increased over the course of continuous rater training from t0 to t2 in all 13 capacity dimensions. Learning effects of the raters can be assumed. It can also be assumed that along with the rater training, rating rules have been internalized over time. In Switzerland very specific quality criteria and in detail description of the rating rules are used in psychological assessments [54], which might have also contributed to higher inter-rater reliabilities over the course until t2. Concluding, our study shows that collegial exchange and rater trainings for measurement of capacity impairments are useful to increase inter-rater agreements.

Second, there were no significant changes in the participants' capacity (impairment) levels over the course of a long-term vocational training program. In most capacities there was even a tendency for increased impairment. Explanations could be the either the findings show a real tendency to worsen the capacities, or that raters become more critical in their assessment over the time, or that the work required of participants to the vocational training program becomes more and more complicated over time. Possibly the raters get a more detailed knowledge of the participant's capacities and impairments due to several observation appointments. Another study [55] has discovered similar findings on increasing capacity impairments during a five-week occupational therapy treatment. Depending on the changes over the course of the vocational training (e.g. changing working settings, further developments in qualification stages or the participant), capacity impairments may be assessed with different reference contexts at different times. Furthermore, observation data (e.g. concrete behaviors) can only be accumulated

Table 1 Professional background of raters

	N	%
t0	24	33
Consultant	48	67
Job Attendant		
t1 (3 months)	20	27
Consultant	54	73
Job Attendant		
t2 (6–9 months)	15	30
Consultant	41	70

over the course, but are not all known in the beginning (t0). For example, a person in training for a salesperson may appear unimpaired in group capacity when asked before the training (t0), but during the training course impairments may become visible (e.g. if the person is on a work placement in a company and can be observed to produce conflicts in the team). When impairments become visible over the course, it is possible that a capacity impairment can be rated at t1 or t2 (in contrast to t0 when no impairment was observable).

Linden and Noack [55] interpreted the effects as a changed rater assessment. Similarly, in our present study it can be assumed that the raters got a more sophisticated assessment of the training participants over several months and thus evaluated more detailed and critically.

Most capacity impairments were low to moderate (rating 1–2 in a range of 0–4). Due to the setting, it can be assumed that the sample of vocational trainees was homogeneous in that there were moderate impairments (which build the basis for treatments like vocational trainings), but not total impairments (in this case there would be no basis for a positive prognosis of a vocational training).

Limitations

A limitation for the interpretation of the results was the small number of 14 (t2) to 19 (t0) training reports for the calculation of the inter-rater reliabilities and that (due to ratings from mixed professions) no statement can be made about the handling of the Mini-ICF of different professional groups. Another limitation may be a rather low power for the analysis of variance with

Table 2 Spearman correlations as measures of agreement between the assessments by master social professionals and consultants for the 13 dimensions of the Mini-ICF-APP at the three different points of measurement

Mini-ICF-APP dimension	r_{t0} *T0* (*N* = 19)	*p*	r_{t1} *T1* (*N* = 17)	*p*	r_{t2} *T2* (*N* = 14)	*p*
Adherence to regulations	0.57*	0.011	0.36	0.162	0.93**	< 0.001
Planning and structuring of tasks	0.57*	0.011	0.39	0.127	0.66**	0.010
Flexibility	0.55*	0.015	0.49*	0.047	0.76**	0.002
Applying expertise	0.24	0.319	0.21	0.413	0.70**	0.005
Capacity to judge and decide	0.26	0.280	0.31	0.222	0.66**	0.010
Endurance	0.10	0.684	0.61**	0.010	0.68**	0.008
Assertiveness	0.50*	0.029	0.64**	0.006	0.76**	0.002
Contacts with others	0.30	0.209	0.64**	0.005	0.55*	0.043
Teamwork capacity	0.66**	0.002	0.69**	0.002	0.73**	0.003
Self-care	0.41	0.084	0.44	0.091	0.73**	0.003
Mobility	0.01	0.959	0.10	0.707	0.97**	< 0.001
Proactivity	0.03	0.910	0.38	0.130	0.90**	< 0.001
Familiar and intimate relationships	0.60*	0.012	0.72**	0.002	0.70**	0.008

*$p = < 0.05$, **$p = < 0.01$

repeated measures. Even though analysis of variance is a very robust procedure and overcomes such limitations [56], future studies should target larger samples. However, this is a sign of the normal naturalistic environment in which the investigation was conducted. As this is a naturalistic sample, no sample size was calculated a priori for the calculations. Moreover, some data were missing and replaced by mean values. The use of the mean as replacement value could have implied a levelling of the overall evolution in some of the 13 capacity dimensions.

Further Discussion and Outlook: Meaning of Vocational Trainings

This present study focused on the assessment of capacity impairments over the course in vocational trainings in naturalistic setting. It can therefore not provide data on the efficacy of vocational trainings. The impact of vocational trainings to increase the capacity levels might be evaluated in randomized controlled intervention studies by using stable and standardized reference contexts. Until now, several research studies have shown the usefulness of vocational trainings [44, 45, 46]. With concrete capacity assessments, evaluations can be done even more differentiated and with focus on behavior, activities and capacities – i.e., what is relevant in the concrete work settings (more than the type of illness or symptoms).

Funding Open Access funding enabled and organized by Projekt DEAL. No funds, grants, or other support was received.

Compliance with ethical standards

Conflicts of interest None.

Ethics approval The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals.

Availability of data and material Data are available upon request from the authors.

Financial interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Muschalla B. Different work capacity impairments in patients with different work-anxieties. Int Arch Occup Environ Health. 2016;89:609–19. https://doi.org/10.1007/s00420-015-1099-x.
2. McHugh ML. Interrater reliability: the kappa statistic. Biochemia Medica. 2012;22:276–82. https://doi.org/10.11613/BM.2012.031.
3. Mahoney FI, Barthel D. Functional evaluation: the Barthel Index. Md State Med J. 1965;14:56–61 (**PMID: 14258950**).
4. Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, Aluwahlia S. A children's global assessment scale (CGAS). Arch Gen Psychiatry. 1983;40:1228–31. https://doi.org/10.1001/archpsyc.1983.01790100074010.
5. Hsueh IP, Lee MM, Hsieh CL. Psychometric characteristics of the Barthel activities of daily living index in stroke patients. J Formos Med Assoc. 2001;100(8):526–32 (**PMID: 11678002**).
6. Lundh A, Kowalski J, Sundberg CJ, Gumpert C, Landén M. Children's Global Assessment Scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings. Psychiatry Res. 2010;177:206–10. https://doi.org/10.1016/j.psychres.2010.02.006.
7. Richards SH, Peters TJ, Coast J, Gunnell DJ, Darlow MA, Pounsford J. Inter-rater reliability of the Barthel ADL index: how does a researcher compare to a nurse? Clin Rehabil. 2000;14:72–8. https://doi.org/10.1191/026921500667059345.
8. Lenze EJ, Munin MC, Quear T, Dew MA, Rogers JC, Begley AE, Reynolds CF III. The Pittsburgh Rehabilitation Participation Scale: reliability and validity of a clinician-rated measure of participation in acute rehabilitation. Arch Phys Med Rehabil. 2004;85:380–4. https://doi.org/10.1016/j.apmr.2003.06.001.
9. Grabe HJ, Hartschen V, Welter-Werner E, Thiel A, Freyberger HJ, Kathmann N, Hoff P. Development of the AMDP module for identification of obsessive-compulsive symptoms. Conceptualization and empirical results. Fortschr Neurol Psychiatr. 1998;66:201–6. https://doi.org/10.1055/s-2007-995256.
10. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, Heninger GR, Charney DS. The yale-brown obsessive compulsive scale. I. Development,

use and reliability. Arch Gen Psychiatry. 1989;46:1006–11. https://doi.org/10.1001/archpsyc.1989.01810110048007.
11. Woody SR, Steketee G, Chambless DL. Reliability and validity of the Yale-Brown obsessive-compulsive scale. Behav Res Ther. 1995;33(5):597–605. https://doi.org/10.1016/0005-7967(94)00076-V.
12. Kovacs M (1992) The Children's Depression Inventory (CDI) Manual. Multi-Health Systems, Inc, North Tonawanda, NY
13. Schneider S, Unnewehr S, Margraf J (2009) Diagnostisches Interview bei psychischen Störungen im Kindes- und Jugendalter (Kinder-DIPS). [Diagnostic interview for mental disorders in childhood and adolescence.]. Springer, Heidelberg. ISBN: 3540782109
14. Allgaier AK, Frühe B, Pietsch K, Saravo B, Baethmann M, Schulte-Körne G. Is the Children's Depression Inventory Short version a valid screening tool in pediatric care? A comparison to its full-length version. J Psychosom Res. 2012;73(5):369–74. https://doi.org/10.1016/j.jpsychores.2012.08.016.
15. Schaub D, Brüne M, Jaspen E, Pajonk FG, Bierhoff HW, Juckel G. The illness and everyday living: close interplay of psychopathological syndromes and psychosocial functioning in chronic schizophrenia. Eur Arch Psychiatry Clin Neurosci. 2011;261:85–93. https://doi.org/10.1007/s00406-010-0122-1.
16. Juckel G, Schaub D, Fuchs N, Naumann U, Uhl I, Witthaus H. Validation of the personal and social performance (PSP) scale in a German sample of acutely ill patients with schizophrenia. Schizophr Res. 2008;104:287–93. https://doi.org/10.1016/j.schres.2008.04.037.
17. Federal Ministery of Labour and Social Affairs (2002). *MELBA - Ein Instrument zur beruflichen Rehabilitation und Integration.* [MELBA - A tool for vocational rehabilitation and integration.] Bonn: Bundesministerium für Arbeit und Sozialordnung BMA. http://www.imba.de/documents/einfuehrung.pdf
18. Achterberg T, Wind H, Prinzie P, Frings-Dresen M. Interrater reliability of the „Merkmalprofile zur Eingliederung Leistungsgewandelter und Behinderter in Arbeit" (MELBA) in young disabled adults with psychosocial limitations. Work. 2013;44:491–7. https://doi.org/10.3233/WOR-2012-1363.
19. Fougeyrollas P, Noreau L, Bergeron H, Cloutier R, Dion SA, St-Michel G. Social consequences of long-term impairments and disabilities: conceptual approach and assessment of handicap. Int J Rehabil Res. 1998;21:127–41. https://doi.org/10.1097/00004356-199806000-00002.
20. Noreau L, Desrosiers J, Robichaud L, Fougeyrollas P, Rochette A, Viscogliosi C. Measuring social participation: reliability of the LIFE-H in older adults with disabilities. Disabil Rehabil. 2004;26:346–52. https://doi.org/10.1080/09638280410001658649.
21. Cerletti, M. (2019). Rentenprüfungsverfahren bei psychischen Störungen–eine Kritik. [Pension review procedure for mental disorders-a critique.] *Bulletin des Médecins Suisses*, *100*, 94–96. doi: https://doi.org/10.4414/bms.2019.17295
22. Hammond, S. (2006). Using psychometric tests. In Breakwell, G. M., Hammond, S., Fife-Schaw, C., & Smith, J. A. (Eds.), *Research methods in psychology* (pp. 182–209). London: Sage Publications. https://psycnet.apa.org/record/2006-10335-000
23. Linden, M., & Baron, S. (2005). Das „Mini-ICF-Rating für psychische Störungen (Mini-ICF-APP)". Ein Kurzinstrument zur Beurteilung von Fähigkeitsstörungen bei psychischen Erkrankungen. [The "Mini-ICF Rating for Mental Disorders (Mini-ICF-APP)". A brief instrument for the assessment of incapacity disorders in mental illness.] *Die Rehabilitation*, *44*, 144–151. doi: https://doi.org/10.1055/s-2004-834786
24. WHO (2001). *International Classification of Functioning Disability and Health: ICF*. Genf: World Health Organisation. https://apps.who.int/iris/bitstream/handle/10665/42407/9241545429.pdf;jsessionid=7624F7B0983BA2687BFD7F831A7D4662?sequence=1
25. Linden, M., Baron, S., & Muschalla, B. (2009, 2015). Mini-ICF-Rating für Aktivitäts- und Partizipationsbeeinträchtigungen bei psychischen Erkrankungen - Manual. [Mini-ICF rating for activity and participation impairments in mental illness - Manual.] Bern: Verlag Hans Huber. https://www.testzentrale.de/shop/mini-icf-rating-fuer-aktivitaets-und-partizipationsbeeintraechtigungen-bei-psychischen-erkrankungen.html
26. Balestrieri M, Isola M, Bonn R, Tam T, Vio A, Linden M, Maso E. Validation of the Italian version of Mini-ICF-APP, a short instrument for rating activity and participation restrictions in psychiatric disorders. Epidemiology and Psychiatric Sciences. 2013;22:81–91. https://doi.org/10.1017/S2045796012000480.
27. Molodynski A, Linden M, Juckel G, Yeeles K, Anderson C, Vazquez-Montes M, Burns T. The reliability, validity and applicability of an English language version of the Mini-ICF-APP. Soc Psychiatry Psychiatr Epidemiol. 2013;48:1347–54. https://doi.org/10.1007/s00127-012-0604-8.
28. Wciórka J, Anczewska M, Jahołkowski P, Świtaj P. Psychometric evaluation of the polish version of the MINI-ICF-APP–a concise measure of limitations on activity and restrictions on participation according to the international classification of functioning, disability and health (ICF)–in people with mental disorders. Advances in Psychiatry and Neurology. 2018;27:218–31. https://doi.org/10.5114/ppn.2018.78715.
29. AWMF (2019). *Leitlinie zur Begutachtung psychischer und psychosomatischer Störungen.* [Guideline for the assessment of mental and psychosomatic disorders.] Berlin: Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften AWMF. https://doi.org/10.1007/s00115-020-00982-1
30. Muschalla B. A concept of psychological work capacity demands – first evaluation in rehabilitation patients with and without mental disorders. Work. 2018;59:375–86. https://doi.org/10.3233/WOR-182691.
31. Linden, M., Muschalla, B., Baron, S., & Ostholt-Corsten, M. (2018). Exploration mittels Mini-ICF-APP - Arbeits- und Leistungsfähigkeitsbeeinträchtigungen bei psychischen Erkrankungen - Ein Fallbeispiel. [Exploration using Mini-ICF-APP - Work and performance impairments in mental illness - A case study.] Berlin: Deutsche Rentenversicherung. https://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Experten/infos_reha_einrichtu

ngen/klassifikationen/miniICF.pdf?__blob=publicationFile&v=2
32. Gerber, K. (2018). Testgütekriterien im Rahmen der medizinischen Begutachtung am Beispiel der ICF. [Test quality criteria in the context of medical assessment using the ICF as an example.] *SZS Schweizerische Zeitschrift für Sozialversicherung und berufliche Vorsorge*. ISSN: 0255–9072
33. Barth J, de Boer W, Busse J, Hoving J, Kedzia S, Couban R, Fischer K, von Allmen D, Spanjer J, Kunz R. Inter-rater agreement in evaluation of disability: systematic review of reproducibility studies. BMJ. 2017. https://doi.org/10.1136/bmj.j14.
34. Kunz R, von Allmen DY, Marelli R, Hoffmann-Richter U, Jeger J, Mager R, Colomb E, Schaad HJ, Bachmann M, Vogel N, Busse JW, Eichhorn M, Banziger O, Zumbrunn T, de Boer WEL, Fischer K. The reproducibility of psychiatric evaluations of work disability: two reliability and agreement studies. BMC Psychiatry. 2019;19:205. https://doi.org/10.1186/s12888-019-2171-y.
35. Lesgold AM. The nature and methods of learning by doing. Am Psychol. 2001;56(11):964. https://doi.org/10.1037/0003-066X.56.11.964.
36. Gaynor M, Seider H, Vogt WB. The volume-outcome effect, scale economies and learning-by-doing. American Economic Review. 2005;95(2):243–7. https://doi.org/10.1257/000282805774670329.
37. Nielsen K. On learning psychotherapy from clients. Nordic Psychology. 2008;60(3):163–82. https://doi.org/10.1027/1901-2276.60.3.163.
38. Dush DM, Hirt ML, Schroeder HE. Self-statement modification in the treatment of child behavior disorders: A meta-analysis. Psychol Bull. 1989;106(1):97. https://doi.org/10.1037/0033-2909.106.1.97.
39. Lyons LC, Woods PJ. The efficacy of rational-emotive therapy: A quantitative review of the outcome research. Clin Psychol Rev. 1991;11(4):357–69. https://doi.org/10.1016/0272-7358(91)90113-9.
40. Dauwerse L, Stolper M, Widdershoven G, Molewijk B. Prevalence and characteristics of moral case deliberation in Dutch health care. Med Health Care Philos. 2014;17(3):365–75. https://doi.org/10.1007/s11019-013-9537-6.
41. Warshaw MG, Dyck I, Allsworth J, Stout RL, Keller MB. Maintaining reliability in a long-term psychiatric study: an ongoing inter-rater reliability monitoring program using the longitudinal interval follow-up evaluation. J Psychiatr Res. 2001;35(5):297–305. https://doi.org/10.1016/S0022-3956(01)00030-9.
42. Belur J, Tompson L, Thornton A, Simon M. Interrater reliability in systematic review methodology: exploring variation in coder decision-making. Sociological methods & research. 2018;13:9372. https://doi.org/10.1177/0049124118799372.
43. Govaerts MJ, Schuwirth LWT, Van der Vleuten CPM, Muijtjens AMM. Workplace-based assessment: effects of rater expertise. Adv Health Sci Educ. 2011;16(2):151–65. https://doi.org/10.1007/s10459-010-9250-7.
44. Lysaker PH, Davis LW, Bryson GJ, Bell MD. Effects of cognitive behavioral therapy on work outcomes in vocational rehabilitation for participants with schizophrenia spectrum disorders. Schizophr Res. 2009;107:186–91. https://doi.org/10.1016/j.schres.2008.10.018.
45. Bejerholm U, Larsson ME, Johanson S. Supported employment adapted for people with affective disorders—A randomized controlled trial. J Affect Disord. 2017;207:212–20. https://doi.org/10.1016/j.jad.2016.08.028.
46. Watzke S, Galvao A, Brieger P. Vocational rehabilitation for subjects with severe mental illnesses in Germany. Soc Psychiatry Psychiatr Epidemiol. 2009;44:523–31. https://doi.org/10.1007/s00127-008-0466-2.
47. Wallace CJ, Tauber R. Rehab rounds: Supplementing supported employment with workplace skills training. Psychiatr Serv. 2004;55:513–5. https://doi.org/10.1176/appi.ps.55.5.513.
48. Berglund E, Anderzén I, Andersén Å, Carlsson L, Gustavsson C, Wallman T, Lytsy P. Multidisciplinary intervention and Acceptance and Commitment Therapy for return-to-work and increased employability among patients with mental illness and/or chronic pain: A randomized controlled trial. Int J Environ Res Public Health. 2018;15:2424. https://doi.org/10.3390/ijerph15112424.
49. Michon HWC, van Weeghel J, Kroon H, Schene AH (2005) Person-related predictors of employment outcomes after participation in psychiatric vocational rehabilitation programmes. Soc Psychiatry Psychiatr Epidemiol 40(5):408–416
50. Bond GR, Drake RE, Luciano A (2015) Employment and educational outcomes in early intervention programmes for early psychosis: a systematic review. Epidemiol Psychiatr Sci 24(5):446–457
51. Donker-Cools BH, Daams JG, Wind H, Frings-Dresen MH. Effective return-to-work interventions after acquired brain injury: a systematic review. Brain Inj. 2016;30:113–31. https://doi.org/10.3109/02699052.2015.1090014.
52. Habermeyer B, Kaiser S, Kawohl W, Seifritz E. Assement of incapacity to work and the Mini-ICF-APP. Neuropsychiatrie. 2017;31:182–6. https://doi.org/10.1007/s40211-017-0246-x.
53. Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation and sample size requirements. Phys Ther 85(3):257–268. https://doi.org/10.1093/ptj/85.3.257.
54. Ebner G, Colomb E, Mager R, Marelli R, Rota F (2016) Qualitätsleitlinien für versicherungspsychiatrische Gutachten. [Quality guidelines for insurance psychiatric reports.]. Schweizer Gesellschaft für Versicherungspsychiatrie SGVP, Schweizerische Gesellschaft für Psychiatrie und Psychotherapie SGPP, Basel. https://www.sgvp.ch/download/385/
55. Linden M, Noack N (2017) Veränderungen in der Beurteilung des (Arbeits-) Fähigkeitsprofils psychosomatischer Patienten im Verlauf einer ergotherapeutischen Behandlung. [Changes in the assessment of the (work)ability profile of psychosomatic patients in the course of occupational therapy treatment.] Arbeitsmedizin Sozialmedizin Umweltmedizin 52:271–275. https://www.asu-arbeitsmedizin.com/beurteilung-des-arbeits-faehigkeitsprofils-psychosomatischer-patienten/originalia-veraenderungen
56. Donaldson TS (1968) Robustness of the F-test to errors of both kinds and the correlation between the numerator and

denominator of the F-ratio. J Am Stat Assoc 63(322):660–676. https://doi.org/10.1080/01621459.1968.11009285.