

Appendix 4: Performance of study and study findings (long) [posted as supplied by author]

Author, year	Number of ... Raters / Cases Total / Cases per rater / Raters per case	Instrument	Outcome and Scale	Reliability Statistic	Results	Interrater Reliability
Insurance Setting						
de Kort, 1992(5)	a) Among 5 raters or b) 5 raters vs. decision of government / 180 / 180 / 5	Professional expertise of raters. Records include: <ul style="list-style-type: none"> • Structured presentation of job specification. • Comprehensive medical questionnaire (by patient). • Medical examination form (by doctor). No specific instrument.	Global rating of 'Fit for Work'. Three categories: Fit / doubt / unfit for job	Percentage agreement; kappa. Precision presented.	Agreement _{Fit for work} among panel physicians Any disorder: Agreement 69%; kappa 0.38 (SD 0.06); Musculoskeletal: Agreement 66%; kappa 0.38 (SD 0.11); Mental disorder: Agreement 58%; kappa 0.24 (SD 0.11); Agreement _{Fit for work panel vs government} ^a Any disorder: Agreement 63%; kappa 0.35 (SD 0.07); Musculoskeletal: Agreement 64%; kappa 0.34 (SD 0.14); Mental disorder: Agreement 61%; kappa 0.33 (SD 0.22);	Kappa poor Agreement poor
Dell-Kuster, 2014(4)	Unclear / 3562 / unclear / 2	Professional expertise of raters. No specific instrument.	Global rating of 'Work Ability'. Scale from 100% to 0%	Percentage agreement. No precision.	Overall agreement among experts and physicians - Last job: 51%. - Suitable alternative work: 20%. Experts approved higher level of work ability than did treating physicians.	Agreement poor
Dickmann 2007(11)	22 / 1 / 1 / 22	Professional expertise of raters. No specific instrument.	'Work Ability' in the last job and in suitable alternative work. Three categories: < 3 hrs; 3 to 6 hrs; ≥ 6 hrs.	Narrative in the text or in descriptive tables. No precision.	Frequency _{Work ability last job} <3hrs: 27%; 3 to 6hrs: 36%; ≥ 6 hrs: 37% Frequency _{Work ability alternative work} <3hrs: 18%; 3 to 6hrs: 32%; ≥ 6 hrs: 50%;	Descriptive: Agreement poor
Elder, 1994(13)	35 / 10 / 10 / 35	Professional expertise of raters. No specific instrument.	Approval or decline of application for early retirement due to ill health. Four categories: Accept / reject / other action / no response	Weighted kappa. Precision presented.	Kappa _{Early retirement} 0.24 (SE 0.03) ^b	Kappa poor
Ikezawa, 2010(9)	36 / 3 / 3 / 36	Professional expertise of raters. Standardized questionnaire based on Alberta's current Workers' Compensation Board (WCB)	'Return to work' recommendation. Three categories:	Percentage agreement; kappa. No precision.	RTW recommendations to modified duties: Global agreement _{RTW recommendation} 82.4% Agreement _{RTW fracture} 97%	Agreement good

Author, year	Number of ... Raters / Cases Total / Cases per rater / Raters per case	Instrument	Outcome and Scale	Reliability Statistic	Results	Interrater Reliability
		outcome reporting requirements	Return to previous work / return to modified work / no return to work		Agreement RTW dislocation 94% Agreement RTW chronic back pain 56%	
Ingravallo, 2008(2)	4 / 15 / 15 / 4	Professional expertise of the members of the Medical Commissions. No specific instrument.	a) Decision on disability benefit Percentage disability on a scale from 0% to 100%. b) Three categories on severity of handicap: No handicap / handicap / severe handicap	Percentage agreement; kappa. Precision presented.	Disability decision Agreement Disability benefit range 20% to 53% Kappa Disability benefit range -0.10 [95% CI: -0.28 to 0.07] to 0.35 [95% CI: 0.01 to 0.69] Handicap decision Agreement Severity handicap range 13.3% to 60% Kappa Severity handicap range -0.26 [95% CI: -0.36 to 0.15] to 0.36 [95% CI: 0.039 to 0.68]	Kappa poor Agreement poor
Lax, 2004(7)	2 / 23 / 23 / 2 ^c	Professional expertise of the occupational health center (OHC) examiner and the independent medical expert (IME). No specific instrument.	Agreement among OHC examiner and IME on four disability items: - presence or absence - partial or total - temporary or permanent - apportioned between work and non-work related. Three categories: Full / partial / dis-agreement.	Narrative in the text or in descriptive tables. No precision.	Full agreement Four disability items 4% Partial agreement Four disability items 34% Disagreement Four disability items 78% Direction of experts' disagreement favours insurers.	Descriptive: Agreement poor
Lederer, 1998(8)	103 / 1 / 1 / 103 ^d	General instructions for disability assessments by public health physicians. Professional expertise of raters. No specific instrument.	'Fit for Work' evaluation. Two categories: Yes vs. No.	Narrative in the text or in descriptive tables. No precision.	Frequency Remaining work ability 91.2% Of 94 judgements with limited work ability Frequency Limitations in work performance 86% Frequency Reduction in working hours 56%	Descriptive: Agreement good ^e
Okpaku, 1994(10)	6 / 144 / 48 to 129 / 2 to 6	Professional expertise of mental health workers. Criteria of the Social Security Administration (SSA). No specific instrument.	Approval for Social Security Benefit. SSA: Two categories: Yes vs. No. Mental health workers (Team): Four categories: Yes / maybe / no / undecided.	Narrative in the text or in descriptive tables. No precision.	Team No (20/144) SSA Yes (11/20) Team Yes (56/144) SSA Yes (49/56) Team Maybe (29/144) SSA Yes (23/29) Team Undecided (39/144) SSA Yes (31/39) Agreement Yes / No decisions 40% Direction of mental health workers' disagreement favours insurer.	Descriptive: Poor
Rudbeck, 2011(14)	11 / 8 / 8 / 11	Professional expertise of raters. No specific instrument.	Health related work ability. Four categories:	Weighted kappa. No precision.	Kappa Work ability 0.33	Kappa poor

Author, year	Number of ... Raters / Cases Total / Cases per rater / Raters per case	Instrument	Outcome and Scale	Reliability Statistic	Results	Interrater Reliability
			Intact / slightly reduced / much reduced / extremely reduced.			
Schellart, 2013(12)	40 / 4 / 4 / 40	Professional expertise of raters. List of Functional Abilities (LFA), with mental abilities, general physical abilities including manual skills, autonomy.	LFA sum score based on four scales 'Mental Abilities', 'General Physical Ability', 'Autonomy' and 'Working Hours'. No information about ability and autonomy scales.	Intraclass correlation. Precision presented.	Pre-intervention ICC _{LFA sum score} 0.64 [95% CI 0.32 to 0.83] Pre-control ICC _{LFA sum score} 0.65 [95% CI 0.33 to 0.84] Post-intervention ICC _{LFA sum score} 0.69 [95% CI 0.37 to 0.86] Post-control ICC _{LFA sum score} 0.54 [95% CI 0.21 to 0.76]	ICC good
Schreuder, 2012(15)	5 / 132 / 132 / 5	Professional expertise of raters. No specific instrument.	Readiness and ability of employee to return to work. Two categories: High vs. low	Percentage agreement. Mean of kappa values of 10 pairs of raters. No precision.	Any disorder Agreement _{Ability} 57%; kappa _{Ability} 0.14 Agreement _{Willingness} 63%; kappa _{Willingness} 0.25 Mental disorders Agreement _{Ability} 55%; kappa _{Ability} 0.05 Agreement _{Willingness} 58%; kappa _{Willingness} 0.18 Musculoskeletal disorders Agreement _{Ability} 59%; kappa _{Ability} 0.22 Agreement _{Willingness} 68%; kappa _{Willingness} 0.38	Kappa poor Agreement poor
Slebus, 2010(16)	51 / 5 / 5 / 25 to 26	Professional expertise of raters. Checklist for work ability.	Global rating of work ability. Percentage work ability on a scale from 0% ('no work ability') to 100% ('work ability as before the depressive episode')	Intraclass correlation. No precision.	ICC _{Work ability} 0.64	ICC good
Spanjer, 2008(18)	12 / 12 / 12 / 12	Professional expertise of raters. DASI-interview; Functional Information System (FIS); Mental Ability List (MAL).	Reduction in working hours. Hours per day.	Kappa and precision. ^f	Kappa _{Reduction in working hours} 0.80 (0.55–0.94)	Kappa excellent
Spanjer, 2009(3)	27 / 30 / 30 / 27	Professional expertise of raters. Functional Ability List (FAL).	Reduction in working hours. Hours per day.	Kappa and precision. ^f	Kappa _{Reduction in working hours} 0.86 (0.75–0.94)	Kappa excellent
Spanjer, 2010(6)	16 / 62 / 29 / 2	Control: Professional expertise of raters.	Reduction in working hours. Hours per day.	Intraclass correlation. ^f	ICC _{Reduction in working hours} 0.53	ICC fair

Author, year	Number of ... Raters / Cases Total / Cases per rater / Raters per case	Instrument	Outcome and Scale	Reliability Statistic	Results	Interrater Reliability
		Functional Ability List (FAL). Intervention: DASI-interview added.				
Research setting						
Berns, 2007(19)	Unclear / 49 / unclear / 2	Multidimensional Scale of Independent Functioning, here: Bipolar Disorders. Semi-structured interview instrument. Professional expertise ('psychiatrists').	Global rating about functioning within the work environment, reflecting ratings on role, level of support and performance. 7-item Likert scale (1 = normal functioning; 7 = total disability).	Intraclass correlation. No precision.	ICC _{Work functioning} 0.86 Subdomains - ICC _{Role} 0.87 - ICC _{Support} 0.99 - ICC _{Performance} 0.99	ICC excellent
Chopra, 2002(20)	2 / 20 / 20 / 2	Semistructured interview based on the ICIDH-2 checklist, with a single item 'remunerative employment'. Professional expertise (clinicians).	Global rating for remunerative employment. 5-item scale (no / mild / moderate / severe / complete or extreme problem).	Kappa. No precision.	Weighted kappa _{Remunerative employment} 0.62	Kappa good
Daradkeh, 1994(21)	2 / 42 / 42 / 2	Axis V (disabilities) of the ICD 10 with a single item 'Occupational Functioning'. Professional expertise (psychiatrists).	Occupational functioning. 6-point scale from "no dysfunction" to "maximum dysfunction".	Weighted kappa. No precision.	Kappa _{Occupational functioning} 0.53	Kappa fair
Hannula, 2006(22)	4 / 39 / 39 / 4	Occupational Functioning Scale. One-item rating scale, anchored every ten points with a description of a patient's state at that point. Professional expertise (3 psychiatrists, 1 psychologist).	Occupational functioning. Scale from 100 to 0, anchored at every 10 points, higher values indicate better functioning.	Intraclass correlation. Precision presented.	Overall agreement between 4 raters ICC _{Occupational functioning} 0.91 (0.86-0.95)	ICC excellent
Hill, 1989(23)	3 / 21 / 21 / 3	Adult Personality Functioning Assessment where work is 1 out of 6 domains of functioning. Assessment based on a standardised interview developed by the authors. Professional expertise (2	Dysfunctioning in work as a social role. 6-point scale from 0 to 5, higher values indicate worse functioning.	Intraclass correlation. Precision presented.	ICC _{Work functioning} 0.76 [95% CI, lower limit: 0.61] Pairs of raters: - Raters 1 vs. 2: ICC _{Work functioning} 0.74 - Raters 1 vs. 3: ICC _{Work functioning} 0.77 - Raters 2 vs. 3: ICC _{Work functioning} 0.77	ICC excellent

Author, year	Number of ... Raters / Cases Total / Cases per rater / Raters per case	Instrument	Outcome and Scale	Reliability Statistic	Results	Interrater Reliability
		psychiatrists, 1 researcher).				
Mundo, 2010(24)	18 / 180 / 11 to 41 / 2	Kennedy Axis V multi-dimensional scale with work as 1 of 7 subscales of functioning. Professional expertise (psychiatrists, psychologists, psychotherapists, other mental health professionals).	Occupational skills Scale from 100 to 0, anchored at every 10 points, higher values indicate better occupational skills.	Intraclass correlation. No precision presented.	ICC _{Occupational skills} 0.75	ICC excellent
Nozu, 1995(25)	3 / 20 / 20 / 3	Work-Personality Insufficiency (WPI) Rating Scale. Professional expertise (experts in occupational rehabilitation).	Employability	Intraclass correlation. No precision presented.	ICC _{Employability} 0.88	ICC excellent

Abbreviations: IRR = Interrater Reliability; RTW = Return To Work; ICC = Intraclass Coefficient.

^a This study compared two different perspectives on the client in one sub-analysis.

^b Kappa analysis restricted to 'accept'/'reject' options.

^c In this study, 23 patients were assessed. Some were examined more than once leading to 27 examinations.

^d The numbers were adjusted for the available information in the study.

^e The importance of outcomes for decision making (remaining work ability > limitations in work performance > reduction in working hours) determined the global judgement.

^f Raw data provided by the author allowed to calculate kappa and ICC on working hours.

1. Edwards W. The theory of decision making. *Psychological bulletin*. 1954;51(4):380-417.
2. Ingravallo F, Vignatelli L, Brini M, Brugaletta C, Franceschini C, Lugaresi F, et al. Medico-legal assessment of disability in narcolepsy: an interobserver reliability study. *Journal of sleep research*. 2008;17(1):111-9.
3. Spanjer J, Krol B, Popping R, Groothoff JW, Brouwer S. Disability assessment interview: the role of detailed information on functioning in addition to medical history-taking. *Journal of rehabilitation medicine*. 2009;41(4):267-72.
4. Dell-Kuster S, Lauper S, Koehler J, Zwimpfer J, Altermatt B, Zwimpfer T, et al. Assessing work ability--a cross-sectional study of interrater agreement between disability claimants, treating physicians, and medical experts. *Scandinavian journal of work, environment & health*. 2014;40(5):493-501.
5. de Kort WL, Uiterweer HW, van Dijk FJ. Agreement on medical fitness for a job. *Scandinavian journal of work, environment & health*. 1992;18(4):246-51.
6. Spanjer J, Krol B, Brouwer S, Popping R, Groothoff JW, van der Klink JJ. Reliability and validity of the Disability Assessment Structured Interview (DASI): a tool for assessing functional limitations in claimants. *Journal of occupational rehabilitation*. 2010;20(1):33-40.
7. Lax MB, Manetti FA, Klein RA. Medical evaluation of work-related illness: evaluations by a treating occupational medicine specialist and by independent medical examiners compared. *Int J Occup Environ Health*. 2004;10(1):1-12.
8. Lederer P, Pfaff G, Walter K, Weihrauch M, Weber A. [Quality circles in expert assessment as an instrument in quality management]. *Gesundheitswesen*. 1998;60(7):415-9.
9. Ikezawa Y, Battie MC, Beach J, Gross D. Do clinicians working within the same context make consistent return-to-work recommendations? *Journal of occupational rehabilitation*. 2010;20(3):367-77.
10. Okpaku SO, Sibulkin AE, Schenzler C. Disability determinations for adults with mental disorders: Social Security Administration vs independent judgments. *American journal of public health*. 1994;84(11):1791-5.
11. Dickmann JR, Broocks A. [Psychiatric expert opinion in case of early retirement--how reliable?]. *Fortschr Neurol Psychiatr*. 2007;75(7):397-401.
12. Schellart AJ, Zwerver F, Anema JR, Van der Beek AJ. The influence of applying insurance medicine guidelines for depression on disability assessments. *BMC research notes*. 2013;6:225.
13. Elder AG, Symington IS, Symington EH. Do occupational physicians agree about ill-health retiral? A study of simulated retirement assessments. *Occupational medicine*. 1994;44(5):231-5.
14. Rudbeck M, Fonager K. Agreement between medical expert assessments in social medicine. *Scand J Public Health*. 2011;39(7):766-72.
15. Schreuder JA, Roelen CA, de Boer M, Brouwer S, Groothoff JW. Inter-physician agreement on the readiness of sick-listed employees to return to work. *Disabil Rehabil*. 2012;34(21):1814-9.
16. Slebus FG, Kuijer PP, Willems JH, Frings-Dresen MH, Sluiter JK. Work ability assessment in prolonged depressive illness. *Occupational medicine*. 2010;60(4):307-9.
17. Slebus FG, Kuijer PP, Willems JH, Frings-Dresen MH, Sluiter JK. Work ability in sick-listed patients with major depressive disorder. *Occupational medicine*. 2008;58(7):475-9.
18. Spanjer J, Krol B, Brouwer S, Groothoff JW. Inter-rater reliability in disability assessment based on a semi-structured interview report. *Disabil Rehabil*. 2008;30(24):1885-90.
19. Berns S, Uzelac S, Gonzalez C, Jaeger J. Methodological considerations of measuring disability in bipolar disorder: validity of the Multidimensional Scale of Independent Functioning. *Bipolar disorders*. 2007;9(1-2):3-10.

20. Chopra P, Couper J, Herrman H. The assessment of disability in patients with psychotic disorders: an application of the ICDH-2. *The Australian and New Zealand journal of psychiatry*. 2002;36(1):127-32.
21. Daradkeh TK, Saad A. The reliability and validity of the proposed axis V (disabilities) of ICD-10. *The British journal of psychiatry : the journal of mental science*. 1994;165(5):683-5.
22. Hannula JA, Lahtela K, Jarvikoski A, Salminen JK, Makela P. Occupational Functioning Scale (OFS)--an instrument for assessment of work ability in psychiatric disorders. *Nordic journal of psychiatry*. 2006;60(5):372-8.
23. Hill J, Harrington R, Fudge H, Rutter M, Pickles A. Adult personality functioning assessment (APFA). An investigator-based standardised interview. *The British journal of psychiatry : the journal of mental science*. 1989;155:24-35.
24. Mundo E, Bonalume L, Del Corno F, Madeddu F, Lang M. L'applicazione dell'Asse V di Kennedy a un campione clinico italiano. *Rivista di Psichiatria*. 2010;45(4).
25. Nozu M. [Evaluating work-personality insufficiency of schizophrenic patients; an assessment of employability in psychiatric rehabilitation]. *Seishin shinkeigaku zasshi = Psychiatria et neurologia Japonica*. 1995;97(4):217-38.