



Universiteit
Leiden
The Netherlands

Computerized adaptive testing in Dutch mental health care: a new tool to assess depression and anxiety

Slok-Flens, G.

Citation

Slok-Flens, G. (2022, October 5). *Computerized adaptive testing in Dutch mental health care: a new tool to assess depression and anxiety*. Retrieved from <https://hdl.handle.net/1887/3466118>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3466118>

Note: To cite this publication please use the final published version (if applicable).

References

- Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The reliability of psychiatric diagnosis revisited: The clinician's guide to improve the reliability of psychiatric diagnosis. *Psychiatry (Edgmont)*, 3(1), 41.
- Agresti, A., & Kateri, M. (2011). Categorical data analysis. In *International encyclopedia of statistical science* (pp. 206-208). Berlin, Germany: Springer.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1974). *Standards for educational & psychological tests*.
- Amtmann, D., Kim, J., Chung, H., Bamer, A. M., Askew, R. L., Wu, S., ... & Johnson, K. L. (2014). Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. *Rehabilitation psychology*, 59(2), 220.
- Arrindell, W. A. M., & Ettema, H. (1981). Dimensionele structuur, betrouwbaarheid en validiteit van de Nederlandse bewerking van de Symptom Checklist (SCL-90): Gegevens gebaseerd op een fobisch en een "normale" populatie. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*.
- Babcock, B., & Weiss, D. J. (2013). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement?. *Journal of Computerized Adaptive Testing*, 1, 1-18.
- Bass, M., Morris, S., & Neapolitan, R. (2015). Utilizing multidimensional computer adaptive testing to mitigate burden with patient reported outcomes. In *AMIA Annual Symposium Proceedings* (Vol. 2015, p. 320). American Medical Informatics Association.
- Baxter, A. J., Scott, K. M., Vos, T., & Whiteford, H. A. (2013). Global prevalence of anxiety disorders: a systematic review and meta-regression. *Psychological medicine*, 43(5), 897-910.
- Bechger, T. M., & Maris, G. (2015). A statistical test for differential item pair functioning. *Psychometrika*, 80(2), 317-340.

- Beck, A., & Beamesderfer, A. (1974). Assessment of depression: The depression inventory. In P. Pichot (Ed.), *Psychological measurements in psychopharmacology: Modern problems in pharmacopsychiatry* (Vol. 7, pp. 151–169). Oxford, England: Karger.
- Becker, J., Fliege, H., Kocalevent, R. D., Bjorner, J. B., Rose, M., Walter, O. B., & Klapp, B. F. (2008). Functioning and validity of A Computerized Adaptive Test to measure anxiety (A-CAT). *Depression and anxiety*, 25(12), E182-E194.
- Bedford, A. (1997). On Clark-Watson's tripartite model of anxiety and depression. *Psychological Reports*, 80(1), 125-126.
- Beleckas, C. M., Prather, H., Guattery, J., Wright, M., Kelly, M., & Calfee, R. P. (2018). Anxiety in the orthopedic patient: using PROMIS to assess mental health. *Quality of Life Research*, 27(9), 2275-2282.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, 107(2), 238.
- Bernstein, I. H., & Nunnally, J. C. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Bevans, K. B., Gardner, W., Pajer, K. A., Becker, B., Carle, A., Tucker, C. A., & Forrest, C. B. (2018). Psychometric evaluation of the PROMIS® pediatric psychological and physical stress experiences measures. *Journal of pediatric psychology*, 43(6), 678-692.
- Bijl, R. V., Ravelli, A., & van ZESSEN, G. (1998). Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Social psychiatry and psychiatric epidemiology*, 33(12), 587-595.
- Bijl, R. V., Van Zessen, G., Ravelli, A., De Rijk, C., & Langendoen, Y. (1998). The Netherlands mental health survey and incidence study (NEMESIS): objectives and design. *Social psychiatry and psychiatric epidemiology*, 33(12), 581-586.
- Bock, D. (1997). A brief history of item theory. *Educational measurement: issues and practice*, 16(4), 21-33.
- Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.
- Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in human behavior*, 23(1), 463-477.
- Borsboom, D. (2006). When does measurement invariance matter?. *Medical care*, 44(11), S176-S181.

- Brouwer, D., Meijer, R. R., & Zevalkink, J. (2013). Measuring individual significant change on the Beck Depression Inventory-II through IRT-based statistics. *Psychotherapy Research*, 23(5), 489-501.
- Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Carlier, I. V., Meuldijs, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *Journal of evaluation in clinical practice*, 18(1), 104-110.
- Carlier, I. V., Kovács, V., van Noorden, M. S., van der Feltz-Cornelis, C., Mooij, N., Schulte-van Maaren, Y. W., ... & Giltay, E. J. (2017). Evaluating the responsiveness to therapeutic change with routine outcome monitoring: A Comparison of the Symptom Questionnaire-48 (SQ-48) with the Brief Symptom Inventory (BSI) and the Outcome Questionnaire-45 (OQ-45). *Clinical Psychology & Psychotherapy*, 24(1), 61-71.
- Carlier, I., Schulte-Van Maaren, Y., Wardenaar, K., Giltay, E., Van Noorden, M., Vergeer, P., & Zitman, F. (2012). Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders. *Psychiatry research*, 200(2-3), 904-910.
- Carlier, I. V., van Eeden, W. A., de Jong, K., Giltay, E. J., van Noorden, M. S., van der Feltz-Cornelis, C., ... & van Hemert, A. M. (2019). Testing for response shift in treatment evaluation of change in self-reported psychopathology amongst secondary psychiatric care outpatients. *International Journal of Methods in Psychiatric Research*, 28(3), e1785.
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., ... & PROMIS Cooperative Group. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical epidemiology*, 63(11), 1179-1194.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., ... & Rose, M. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical care*, 45(5 Suppl 1), S3.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of statistical Software*, 48(1), 1-29.
- Chalmers, R. P. (2015). *mirtCAT: Computerized adaptive testing with multidimensional item response theory*. Retrieved from <http://CRAN.R-project.org/package=mirtCAT>

- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological methods & research*, 36(4), 462-494.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling: a multidisciplinary journal*, 9(2), 233-255.
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of statistical software*, 39(8), 1.
- Choi, S. W., Grady, M. W., & Dodd, B. G. (2011). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71(1), 37-53.
- Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of abnormal psychology*, 100(3), 316-336.
- Clarke, D. M., & Kissane, D. W. (2002). Demoralization: its phenomenology and importance. *Australian & New Zealand Journal of Psychiatry*, 36(6), 733-742.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Cook, T. D., & Campbell, D. T. (1959). *Quasi-experiments: Design and analysis issues for field settings*. Rand McNally.
- Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings* (Vol. 351). Boston: Houghton Mifflin.
- Cook, K. F., O'Malley, K. J., & Roddey, T. S. (2005). Dynamic assessment of health outcomes: time to let the CAT out of the bag?. *Health services research*, 40(5 Pt 2), 1694-1711.
- Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical care*, 44, S115-S123.
- Crawford, J. R., & Garthwaite, P. H. (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks. *The Clinical Neuropsychologist*, 23(2), 193-204.

- Crins, M. H., Roorda, L. D., Smits, N., De Vet, H. C., Westhovens, R., Cella, D., ... & Terwee, C. B. (2015). Calibration and validation of the Dutch-Flemish PROMIS pain interference item bank in patients with chronic pain. *PloS one*, 10(7), e0134094.
- Crins, M. H., Terwee, C. B., Klausch, T., Smits, N., de Vet, H. C., Westhovens, R., ... & Roorda, L. D. (2017). The Dutch–Flemish PROMIS Physical Function item bank exhibited strong psychometric properties in patients with chronic pain. *Journal of clinical epidemiology*, 87, 47-58.
- Crins, M. H. P., Roorda, L. D., Smits, N., De Vet, H. C. W., Westhovens, R., Cella, D., ... & Terwee, C. B. (2016). Calibration of the Dutch-Flemish PROMIS Pain Behavior item bank in patients with chronic pain. *European Journal of Pain*, 20(2), 284-296.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.
- de Beurs, E., Barendregt, M., Flens, G., van Dijk, E., Huijbrechts, I., & Meering, W. J. (2012). Equivalentie in responsiviteit van veel gebruikte zelfrapportage meetinstrumenten in de geestelijke gezondheidszorg [Equivalence in responsiveness of commonly used self-report questionnaires in mental health]. *Maandblad voor de Geestelijke Volksgezondheid*, 67, 259-264.
- de Beurs, E., Barendregt, M., & Warmerdam, L. (2017). *Behandeluitkomsten: Bron voor Kwaliteitsbeleid in de GGZ*. Boom Uitgevers Amsterdam.
- de Beurs, E., Carlier, I. V., & van Hemert, A. M. (2019). Approaches to denote treatment outcome: Clinical significance and clinical global impression compared. *International journal of methods in psychiatric research*, 28(4), e1797.
- de Beurs, E., den Hollander-Gijsman, M. E., Helmich, S., & Zitman, F. G. (2007). The tripartite model for assessing symptoms of anxiety and depression: Psychometrics of the Dutch version of the mood and anxiety symptoms questionnaire. *Behaviour Research and Therapy*, 45(7), 1609-1617.
- de Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., Van der Wee, N. J., Giltay, E. J., van Noorden, M. S., ... & Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical psychology & psychotherapy*, 18(1), 1-12.
- de Beurs, E., Flens, G., & Williams, G. (2019). Meetresultaten interpreteren in de klinische psychologie: een aantal voorstellen. *De psycholoog*, 54, 10 – 23.

- de Beurs, E., Warmerdam, E. H., Oudejans, S. C. C., Spits, M., Dingemanse, P., De Graaf, S. D. D., ... & van Son, G. E. (2018). Treatment outcome, duration, and costs: A comparison of performance indicators using data from eight mental health care providers in the Netherlands. *Administration and Policy in Mental Health and Mental Health Services Research*, 45(2), 212-223.
- de Beurs, E. & Zitman, F. G. (2005). De Brief Symptom Inventory (BSI): De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90 [The Brief Symptom Inventory: Reliability and validity of a handy alternative for the SCL-90]. *Maandblad Geestelijke Volksgezondheid*, 61, 120-141.
- de Graaf, R., Ten Have, M., van Gool, C., & van Dorsselaer, S. (2012). Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Social psychiatry and psychiatric epidemiology*, 47(2), 203-213.
- de Graaf, R. O. N., Ten Have, M., & van Dorsselaer, S. (2010). The Netherlands mental health survey and incidence study-2 (NEMESIS-2): Design and methods. *International journal of methods in psychiatric research*, 19(3), 125-141.
- de Jong, K., Conijn, J. M., Gallagher, R. A., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, 102002.
- de Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E., Spinhoven, P., & Heiser, W. J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, 14(4), 288-301.
- de Vet, H. C., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: a practical guide*. Cambridge university press.
- Deng, G., Jiang, C., & Li, Y. X. (2012). Clinical utility of the mood and anxiety symptom questionnaire in a Chinese sample of patients with pancreatic cancer. *Gastroenterology Nursing*, 35(3), 193-198.
- Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacol Bull*, 9, 13-28.
- Dieperink, C. J., Pijl, Y. J., Mulder, N., Van Os, J., & Drukker, M. (2008). Langdurig zorgafhankelijke patiënten in de ggz: samenhang met verstedelijking. *Tijdschrift voor Psychiatrie*, 761-769.

- Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public opinion quarterly*, 57(3), 289-304.
- Drenth, P. J. D., & Sijtsma, K. (2005). *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassingen*. Bohn Stafleu Van Loghum.
- Elsman, E. B., Flens, G., de Beurs, E., Roorda, L. D., & Terwee, C. B. (2022). Towards standardization of measuring anxiety and depression: Differential item functioning for language and Dutch reference values of PROMIS item banks. *Plos one*, 17(8), e0273287.
- Embretson, S., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Flens, G., Smits, N., Carlier, I., van Hemert, A. M., & de Beurs, E. (2016). Simulating computer adaptive testing with the Mood and Anxiety Symptom Questionnaire. *Psychological Assessment*, 28(8), 953-962.
- Flens, G., Smits, N., Terwee, C. B., Dekker, J., Huijbrechts, I., & de Beurs, E. (2017). Development of a computer adaptive test for depression based on the Dutch-Flemish version of the PROMIS item bank. *Evaluation & the health professions*, 40(1), 79-105.
- Flens, G., Smits, N., Terwee, C. B., Dekker, J., Huijbrechts, I., Spinhoven, P., & de Beurs, E. (2019). Development of a computerized adaptive test for anxiety based on the Dutch–Flemish version of the PROMIS item bank. *Assessment*, 26(7), 1362-1374.
- Flens, G., Smits, N., Terwee, C. B., Pijck, L., Spinhoven, P., & de Beurs, E. (2021). Practical significance of longitudinal measurement invariance violations in the Dutch–Flemish PROMIS item Banks for Depression and Anxiety: an illustration with ordered-categorical data. *Assessment*, 28(1), 277-294.
- Fliege, H., Becker, J., Walter, O. B., Bjorner, J. B., Klapp, B. F., & Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT). *Quality of life Research*, 14(10), 2277-2291.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, 9(4), 466 - 491.
- Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: an illustration of longitudinal measurement invariance. *Psychological Assessment*, 25(2), 520–531.

- Forbey, J. D., & Ben-Porath, Y. S. (2007). A comparison of the MMPI-2 restructured clinical (RC) and clinical scales in a substance abuse treatment sample. *Psychological Services*, 4(1), 46-58.
- Forkmann, T., Boecker, M., Norra, C., Eberle, N., Kircher, T., Schauerte, P., ... Wirtz, M. (2009). Development of an item bank for the assessment of depression in persons with mental illnesses and physical diseases using Rasch analysis. *Rehabilitation Psychology*, 54(2), 186-197.
- Forrest, C. B., Devine, J., Bevans, K. B., Becker, B. D., Carle, A. C., Teneralli, R. E., ... & Ravens-Sieberer, U. (2018). Development and psychometric evaluation of the PROMIS Pediatric Life Satisfaction item banks, child-report, and parent-proxy editions. *Quality of Life Research*, 27(1), 217-234.
- Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2018). A tipping point for measurement-based care. *Focus*, 16(3), 341-350.
- Frank, J. D. (1974). Psychotherapy: The restoration of morale. *American Journal of Psychiatry*, 131(3), 271-274.
- Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time . . . or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, 28(11), 1354-1367.
- Fries, J., Rose, M., & Krishnan, E. (2011). The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *The Journal of rheumatology*, 38(8), 1759-1764.
- Fries, J. F., Krishnan, E., Rose, M., Lingala, B., & Bruce, B. (2011). Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis research & therapy*, 13(5), 1-8.
- Gardner, W., Shear, K., Kelleher, K. J., Pajer, K. A., Mammen, O., Buysse, D., & Frank, E. (2004). Computerized adaptive measurement of depression: a simulation study. *BMC psychiatry*, 4(1), 1-11.
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B., & Kupfer, D. J. (2012). Development of a computerized adaptive test for depression. *Archives of general psychiatry*, 69(11), 1104-1112.
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B., & Kupfer, D. J. (2014). Development of the CAT-ANX: a computerized adaptive test for anxiety. *American Journal of Psychiatry*, 171(2), 187-194.

- Gliem J, & Gliem R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. In 2003 Midwest research to practice conference in adult, continuing and community education. Columbus, OH.
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational measurement*, 21(4), 347-360.
- Greenhalgh, J., Gooding, K., Gibbons, E., Dalkin, S., Wright, J., Valderas, J., & Black, N. (2018). How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *Journal of patient-reported outcomes*, 2(1), 1-28.
- Guo, T., Xiang, Y. T., Xiao, L. E., Hu, C. Q., Chiu, H. F., Ungvari, G. S., ... & Wang, G. (2015). Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *American Journal of Psychiatry*, 172(10), 1004-1013.
- Hambleton, R. K. (1988). *Principles and selected applications of Item Response Theory* (3rd ed.). New York: American Council on Education.
- Hare, D. L., Toukhsati, S. R., Johansson, P., & Jaarsma, T. (2014). Depression and cardiovascular disease: a clinical review. *European heart journal*, 35(21), 1365-1372.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The Elements of Statistical Learning: Data mining, inference and prediction*. New York: Springer.
- Haverman, L., Grootenhuis, M. A., Raat, H., van Rossum, M. A., van Dulmen-den Broeder, E., Hoppenbrouwers, K., ... & Terwee, C. B. (2016). Dutch–Flemish translation of nine pediatric item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS)®. *Quality of Life Research*, 25(3), 761-765.
- Hayslett, M. M., & Wildemuth, B. M. (2004). Pixels or pencils? The relative effectiveness of Web-based versus paper surveys. *Library & Information Science Research*, 26(1), 73-93.
- Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, 3(4), 424.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.

- Irwin, D. E., Stucky, B., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J. S., ... & DeWalt, D. A. (2010). An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Quality of Life Research*, 19(4), 595-607.
- Jabrayilov, R., Emons, W. H., de Jong, K., & Sijtsma, K. (2017). Longitudinal measurement invariance of the Dutch Outcome Questionnaire-45 in a clinical sample. *Quality of Life Research*, 26(6), 1473-1481.
- Jabrayilov, R., Emons, W. H., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement*, 40(8), 559-572.
- Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior therapy*, 15(4), 336-352.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Jakob, T., Nagl, M., Gramm, L., Heyduck, K., Farin, E., & Glattacker, M. (2017). Psychometric properties of a German translation of the PROMIS® depression item bank. *Evaluation & the health professions*, 40(1), 106-120.
- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.
- Kaat, A. J., Newcomb, M. E., Ryan, D. T., & Mustanski, B. (2017). Expanding a common metric for depression reporting: linking two scales to PROMIS® depression. *Quality of Life Research*, 26(5), 1119-1128.
- Kaplan, D. (1989). A study of the sampling variability and z-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research*, 24(1), 41-57.
- Kays, K., Gathercoal, K., & Buhrow, W. (2012). Does survey format influence self-disclosure on sensitive question items?. *Computers in Human Behavior*, 28(1), 251-256.
- Keogh, E., & Reidy, J. (2000). Exploring the factor structure of the Mood and Anxiety Symptom Questionnaire (MASQ). *Journal of personality assessment*, 74(1), 106-125.
- Khanna, D., Krishnan, E., Dewitt, E. M., Khanna, P. P., Spiegel, B., & Hays, R. D. (2011). The future of measuring patient-reported outcomes in rheumatology: Patient-Reported Outcomes Measurement Information System (PROMIS). *Arthritis care & research*, 63 3(Suppl 11), S486-S490.

- Kim, S., Moses, T., & Yoo, H. H. (2015). Effectiveness of item response theory (IRT) proficiency estimation methods under adaptive multistage testing. *ETS Research Report Series*, 2015(1), 1-19.
- Kim, S. H., Cohen, A. S., Alagoz, C., & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement*, 44(2), 93-116.
- Kocalevent, R. D., Rose, M., Becker, J., Walter, O. B., Fliege, H., Bjorner, J. B., ... & Klapp, B. F. (2009). An evaluation of patient-reported outcomes found computerized adaptive testing was efficient in assessing stress perception. *Journal of Clinical Epidemiology*, 62(3), 278-287.
- Kose, I. A., & Demirtasli, N. C. (2012). Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test length and sample size. *Procedia-Social and Behavioral Sciences*, 46, 135-140.
- Kroenke, K., Baye, F., & Lourens, S. G. (2019). Comparative responsiveness and minimally important difference of common anxiety measures. *Medical care*, 57(11), 890-897.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9), 606-613.
- Kudel, I., Pona, A., Cox, S., Szoka, N., Tabone, L., & Brode, C. (2019). Psychometric properties of NIH PROMIS® instruments in bariatric surgery candidates. *Health Psychology*, 38(5), 359.
- Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice*. Washington, DC: American Psychological Association.
- Lange, A., Schrieken, B., van de Ven, J., & Blankers, M. (2000). De Korte Klachten Lijst (KKL). [The Short Complaint List]. *Directieve therapie*, 20(4), 181-185.
- Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., ... & Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *European psychiatry*, 12(5), 224-231.
- Lee, S. A., Kim, K. H., & Cho, S. M. (2015). Validation of the Mood and Anxiety Symptom Questionnaire in Korean Adolescents. *Psychiatry investigation*, 12(2), 218-226.
- Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., ... & Kroenke, K. (2019). Implementing measurement-based care in behavioral health: a review. *JAMA psychiatry*, 76(3), 324-335.

- Lewis, C. C., Scott, K., Marti, C. N., Marriott, B. R., Kroenke, K., Putz, J. W., ... & Rutkowski, D. (2015). Implementing measurement-based care (iMBC) for depression in community mental health: a dynamic cluster randomized trial study protocol. *Implementation Science*, 10(1), 1-14.
- Leyro, T. M., Zvolensky, M. J., & Bernstein, A. (2010). Distress tolerance and psychopathological symptoms and disorders: a review of the empirical literature among adults. *Psychological bulletin*, 136(4), 576.
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods*, 48(3), 936-949.
- Lin, A., Yung, A. R., Wigman, J. T., Killackey, E., Baksheev, G., & Wardenaar, K. J. (2014). Validation of a short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ) in adolescents and young adults. *Psychiatry research*, 215(3), 778-783.
- Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological methods*, 22(3), 486.
- Liu, Y., & West, S. G. (2018). Longitudinal measurement non-invariance with ordered-categorical indicators: How are the parameters in second-order latent linear growth models affected?. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(5), 762-777.
- Lizzio, V. A., Blanchett, J., Borowsky, P., Meldau, J. E., Verma, N. N., Muh, S., ... & Makhni, E. C. (2019). Feasibility of PROMIS CAT administration in the ambulatory sports medicine clinic with respect to cost and patient compliance: a single-surgeon experience. *Orthopaedic journal of sports medicine*, 7(1), 2325967118821875.
- Lloyd, C. E., Dyer, P. H., & Barnett, A. H. (2000). Prevalence of symptoms of depression and anxiety in a diabetes clinic population. *Diabetic medicine*, 17(3), 198-202.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley Publishing Company.
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy*, 33(3), 335-343.
- Löwe, B., Spitzer, R. L., Williams, J. B., Mussell, M., Schellberg, D., & Kroenke, K. (2008). Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *General hospital psychiatry*, 30(3), 191-199.

- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99.
- Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R. D., Brod, M., ... & Cella, D. (2012). Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Quality of Life Research*, 21(5), 739-746.
- Mancheño, J. J., Cupani, M., Gutiérrez-López, M., Delgado, E., Moraleda, E., Cáceres-Pachón, P., ... & Rojas, Ó. M. L. (2018). Classical test theory and item response theory produced differences on estimation of reliable clinical index in World Health Organization Disability Assessment Schedule 2.0. *Journal of clinical epidemiology*, 103, 51-59.
- Marasini, D., Quatto, P., & Ripamonti, E. (2016). Assessing the inter-rater agreement for ordinal data through weighted indexes. *Statistical methods in medical research*, 25(6), 2611-2633.
- Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D., & Saxena, S. (2012). Depression: A global public health concern. *WHO Department of Mental Health and Substance Abuse*, 1, 6-8.
- Martin-Cook, K., Palmer, L., Thornton, L., Rush, A. J., Tamminga, C. A., & Ibrahim, H. M. (2021). Setting Measurement-Based Care in Motion: Practical Lessons in the Implementation and Integration of Measurement-Based Care in Psychiatry Clinical Practice. *Neuropsychiatric disease and treatment*, 17, 1621-1631.
- Maruyama, G., & Ryan, C. S. (2014). *Research methods in social relations*. John Wiley & Sons.
- McAleavey, A. A., Nordberg, S. S., & Moltu, C. (2021). Initial quantitative development of the Norse Feedback system: a novel clinical feedback system for routine mental healthcare. *Quality of Life Research*, 30(11), 3097-3115.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276-282.
- Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *Journal of Applied Psychology*, 95(4), 728.
- Meade, A. W., & Lautenschlager, G. J. (2004). Same question, different answers: CFA and two IRT approaches to measurement invariance. In *19th Annual Conference of the Society for Industrial and Organizational Psychology*, Chicago.

- Metz, M. J., Veerbeek, M. A., Twisk, J. W., van der Feltz-Cornelis, C. M., de Beurs, E., & Beekman, A. T. (2019). Shared decision-making in mental health care using routine outcome monitoring: results of a cluster randomised-controlled trial. *Social psychiatry and psychiatric epidemiology*, 54(2), 209-219.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton; Berlin, DeGruyter.
- Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1171-1179.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., ... & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*, 63(7), 737-745.
- Muramatsu, K., Kamijima, K., Yoshida, M., Otsubo, T., Miyaoka, H., Muramatsu, Y., & Gejyo, F. (2007). The patient health questionnaire, Japanese version: validity according to the mini-international neuropsychiatric interview-plus. *Psychological reports*, 101(3), 952-960.
- Newton, P., & Shaw, S. (2014). *Validity in educational & psychological assessment*. London: Sage.
- O'Connell, N. S., Dai, L., Jiang, Y., Speiser, J. L., Ward, R., Wei, W., ... & Gebregziabher, M. (2017). Methods for analysis of pre-post data in clinical research: a comparison of five common methods. *Journal of biometrics & biostatistics*, 8(1), 1-8.
- Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, 14(3), 587-598.
- Oosterveld, P., Vorst, H. C., & Smits, N. (2019). Methods for questionnaire design: a taxonomy linking procedures to test goals. *Quality of Life Research*, 28(9), 2501-2512.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied psychological measurement*, 24(1), 50-64.
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289-298.

- Osman, A., Freedenthal, S., Gutierrez, P. M., Wong, J. L., Emmerich, A., & Lozano, G. (2011). The Anxiety Depression Distress Inventory-27 (ADDI-27): a short version of the Mood and Anxiety Symptom Questionnaire-90. *Journal of Clinical Psychology*, 67(6), 591-608.
- Ostini, R., Finkelman, M., & Nering, M. (2015). Selecting Among Polytomous IRT Models. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 285-304). New York: Routledge/Taylor & Francis Group.
- Oude Voshaar, M. A., ten Klooster, P. M., Glas, C. A., Vonkeman, H. E., Taal, E., Krishnan, E., ... & van de Laar, M. A. (2014). Calibration of the PROMIS physical function item bank in Dutch patients with rheumatoid arthritis. *PloS one*, 9(3), e92367.
- Paap, M. C., Born, S., & Braeken, J. (2019). Measurement efficiency for fixed-precision multidimensional computerized adaptive tests: Comparing health measurement and educational testing using example banks. *Applied Psychological Measurement*, 43(1), 68-83.
- Papuga, M. O., Dasilva, C., McIntyre, A., Mitten, D., Kates, S., & Baumhauer, J. F. (2018). Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. *Health Systems*, 7(1), 1-12.
- Paz, S. H., Spritzer, K. L., Morales, L. S., & Hays, R. D. (2013). Evaluation of the patient-reported outcomes information system (PROMIS®) Spanish-language physical functioning items. *Quality of Life Research*, 22(7), 1819-1830.
- Penfield, R. D., & Bergeron, J. M. (2005). Applying a weighted maximum likelihood latent trait estimator to the generalized partial credit model. *Applied Psychological Measurement*, 29(3), 218-233.
- Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., Cella, D., & PROMIS Cooperative Group. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment*, 18(3), 263-283.
- Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Lawrence, S. M., & Daley, D. C. (2016). Validation of the alcohol use item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS®). *Drug and alcohol dependence*, 161, 316-322.
- Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Maihoefer, C. C., & Lawrence, S. M. (2014). Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS®) in a three-month observational study. *Journal of psychiatric research*, 56, 112-119.

- Porter, M. E. (2009). A strategy for health care reform—toward a value-based system. *New England Journal of Medicine*, 361(2), 109-112.
- Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147-1157.
- R Core Team (2014, 2015, 2017, 2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of educational statistics*, 4(3), 207-230.
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied psychological measurement*, 9(4), 401-412.
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., ... & Liu, H. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical care*, 45, S22-S31.
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19-31.
- Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual review of clinical psychology*, 5, 27-48.
- Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of educational Measurement*, 27(2), 133-144.
- Revelle, W. (2013). *psych: Procedures for personality and psychological research*. Evanston, IL: Northwestern University. Retrieved from <http://CRAN.R-project.org/package=psych>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, 17(3), 354.
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, 29(5), 615-620.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses, *Journal of Statistical Software*, 17, 1-25.

- Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36.
- Rush, A. J., & Thase, M. E. (2018). Improving depression outcome by patient-centered medical management. *American Journal of Psychiatry*, 175(12), 1187-1198.
- Samejima, F. (1969). Estimation of latent ability using a pattern of graded responses. *Psychometrika Monograph Supplement*, 34, 100.
- Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 167-180.
- Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In *Innovations in multivariate statistical analysis* (pp. 233-247). Springer, Boston, MA.
- Sawatzky, R., Kwon, J. Y., Barclay, R., Chauhan, C., Frank, L., van den Hout, W. B., ... & Sprangers, M. A. (2021). Implications of response shift for micro-, meso-, and macro-level healthcare decision-making using results of patient-reported outcome measures. *Quality of Life Research*, 30, 3343–3357.
- Schalet, B. D., Pilkonis, P. A., Yu, L., Dodds, N., Johnston, K. L., Yount, S., ... & Cella, D. (2016). Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. *Journal of clinical epidemiology*, 73, 119-127.
- Scott, K., & Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and behavioral practice*, 22(1), 49-59.
- Seidel, J. A., Miller, S. D., & Chow, D. L. (2014). Effect size calculations for the clinician: Methods and comparability. *Psychotherapy Research*, 24(4), 470-484.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., ... & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of clinical psychiatry*, 59(20), 22-33.
- Simms, L. J., Grös, D. F., Watson, D., & O'Hara, M. W. (2008). Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depression and anxiety*, 25(7), E34-E46.
- Simms, L. J., Prisciandaro, J. J., Krueger, R. F., & Goldberg, D. P. (2012). The structure of depression, anxiety and somatic symptoms in primary care. *Psychological medicine*, 42(1), 15-28.

- Singer, S., Das-Munshi, J., & Brähler, E. (2010). Prevalence of mental health conditions in cancer patients in acute care—a meta-analysis. *Annals of oncology*, 21(5), 925-930.
- Smits, N. (2016). On the effect of adding clinical samples to validation studies of patient-reported outcome item banks: a simulation study. *Quality of Life Research*, 25(7), 1635-1644.
- Smits, N., Cuijpers, P., & van Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry research*, 188(1), 147-155.
- Smits, N., Paap, M. C., & Böhnke, J. R. (2018). Some recommendations for developing multidimensional computerized adaptive tests for patient-reported outcomes. *Quality of Life Research*, 27(4), 1055-1063.
- Smits, N., Zitman, F. G., Cuijpers, P., den Hollander-Gijsman, M. E., & Carlier, I. V. (2012). A proof of principle for using adaptive testing in routine Outcome Monitoring: the efficiency of the Mood and Anxiety Symptoms Questionnaire-Anhedonic Depression CAT. *BMC Medical Research Methodology*, 12(1), 1-10.
- Soland, J. (2021). Is measurement noninvariance a threat to inferences drawn from randomized control trials? Evidence from empirical and simulation studies. *Applied Psychological Measurement*, 01466216211013102.
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, 166(10), 1092-1097.
- te Poel, F., Hartmann, T., Baumgartner, S. E., & Tanis, M. (2017). A psychometric evaluation of the Dutch Short Health Anxiety Inventory in the general population. *Psychological assessment*, 29(2), 186.
- Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Ramirez, M., & Kim, G. (2016). Measurement equivalence of the Patient Reported Outcomes Measurement Information System®(PROMIS®) Anxiety short forms in ethnically diverse groups. *Psychological test and assessment modeling*, 58(1), 183-219.
- Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Ramirez, M., & Kim, G. (2016). Psychometric properties and performance of the Patient Reported Outcomes Measurement Information System®(PROMIS®) depression short forms in ethnically diverse groups. *Psychological test and assessment modeling*, 58(1), 141–181.

- Terwee, C. B., Crins, M. H., Roorda, L. D., Cook, K. F., Cella, D., Smits, N., & Schalet, B. D. (2021). International application of PROMIS computerized adaptive tests: US versus country-specific item parameters can be consequential for individual patient scores. *Journal of Clinical Epidemiology*, 134, 1-13.
- Terwee, C. B., Crins, M. H. P., Boers, M., de Vet, H. C. W., & Roorda, L. D. (2019). Validation of two PROMIS item banks for measuring social participation in the Dutch general population. *Quality of Life Research*, 28(1), 211-220.
- Terwee, C. B., Peipert, J. D., Chapman, R., Lai, J. S., Terluin, B., Cella, D., ... & Mokkink, L. B. (2021). Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. *Quality of Life Research*, 30(10), 2729-2754.
- Terwee, C. B., Roorda, L. D., De Vet, H. C. W., Dekker, J., Westhovens, R., Van Leeuwen, J., ... & Boers, M. (2014). Dutch–Flemish translation of 17 item banks from the patient-reported outcomes measurement information system (PROMIS). *Quality of Life Research*, 23(6), 1733-1741.
- Thissen, D. (1991). *MULTILOG user's guide*. Mooresville, IN: Scientific Software.
- Torchiano, M. (2016). Package “effsize” [Software program]. Retrieved from <https://cran.r-project.org/web/packages/effsize/effsize.pdf>
- Valentine, T. R., Weiss, D. M., Jones, J. A., & Andersen, B. L. (2019). Construct validity of PROMIS® Cognitive Function in cancer patients and noncancer controls. *Health Psychology*, 38(5), 351.
- van Bebbber, J., Flens, G., Wigman, J. T., de Beurs, E., Sytema, S., Wunderink, L., & Meijer, R. R. (2018). Application of the Patient-Reported Outcomes Measurement Information System (PROMIS) item parameters for Anxiety and Depression in the Netherlands. *International journal of methods in psychiatric research*, 27(4), e1744.
- van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in psychology*, 6, 1064.
- van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of statistical software*, 20(11), 1-19.
- van der Feltz-Cornelis, C. M., Andrea, H., Kessels, E., Duivenvoorden, H. J., Biemans, H., & Metz, M. (2014). Does routine outcome monitoring have a promising future? An investigation into the use of shared decision-making combined with ROM for patients with a combination of physical and psychiatric symptoms. *Tijdschrift voor psychiatrie*, 56(6), 375-384.

- van Noorden, M. S., Giltay, E. J., den Hollander-Gijsman, M. E., van der Wee, N. J., van Veen, T., & Zitman, F. G. (2010). Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: The Leiden Routine Outcome Monitoring Study. *Journal of affective disorders*, 125(1-3), 116-123.
- van Sonsbeek, M., Hutschemaekers, G. J., Veerman, J. W., Vermulst, A., & Tiemens, B. G. (2021). The results of clinician-focused implementation strategies on uptake and outcome of measurement-based care in general mental health care.
- van Vliet, I. M., & De Beurs, E. (2007). The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders. *Tijdschrift voor psychiatrie*, 49(6), 393-397.
- van Widenfelt, B. M., Treffers, P. D., De Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical child and family psychology review*, 8(2), 135-147.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4-70.
- Verdam, M. G. E., Van Ballegooijen, W., Holtmaat, C. J. M., Knoop, H., Lancee, J., Oort, F. J., ... & Sprangers, M. A. G. (2021). Re-evaluating randomized clinical trials of psychological interventions: Impact of response shift on the interpretation of trial results. *Plos one*, 16(5), e0252035.
- Vilagut, G., Forero, C. G., Adroher, N. D., Olariu, E., Cella, D., Alonso, J., & INSAyD Investigators. (2015). Testing the PROMIS® Depression measures for monitoring depression in a clinical sample outside the US. *Journal of psychiatric research*, 68, 140-150.
- Wagner, L. I., Schink, J., Bass, M., Patel, S., Diaz, M. V., Rothrock, N., ... & Cella, D. (2015). Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. *Cancer*, 121(6), 927-934.
- Wahl, I., Rutsohn, J., Cella, D., Löwe, B., Rose, M., Brähler, E., ... & Schalet, B. (2015). Does anxiety mean the same in English and German language? Evaluation of the psychometric equivalence of the PROMIS® anxiety item bank and its German translation. *Journal of Psychosomatic Research*, 6(78), 629-630.
- Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., & Steinberg, L. (2001). Computerized adaptive testing: A primer. *Quality Life Research*, 10(8), 733-734.

- Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F., & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'Anxiety'(Anxiety-CAT). *Quality of Life Research*, 16(1), 143-155.
- Wang, M. (2016). *Longitudinal differential item functioning detection using bifactor models and the Wald test* (Doctoral dissertation, University of Kansas).
- Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135.
- Wardenaar, K. J., van Veen, T., Giltay, E. J., de Beurs, E., Penninx, B. W., & Zitman, F. G. (2010). Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). *Psychiatry research*, 179(1), 101-106.
- Watson, D., & Clark, L. A. (1991). *The Mood and Anxiety Symptom Questionnaire*. Iowa City, IA: University of Iowa.
- Watson, D., Clark, L. A., Weber, K., Assenheimer, J. S., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *Journal of abnormal Psychology*, 104(1), 15-25.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375.
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child development perspectives*, 4(1), 10-18.
- Windle, C. (1954). Test-retest effect on personality questionnaires. *Educational and Psychological Measurement*, 14(4), 617-633.
- Woods, C. M. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*, 33(1), 42-57.
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014-1045.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of educational measurement*, 30(3), 187-213.

- Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 435-463.
- Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, 39(4), 561-577.