

Assessment, measurement and evaluation within public mental health

Citation for published version (APA):

Hoofs, H. M. (2018). *Assessment, measurement and evaluation within public mental health: methodological and psychometric challenges*. Ipskamp Printing BV.
<https://doi.org/10.26481/dis.20180507hh>

Document status and date:

Published: 01/01/2018

DOI:

[10.26481/dis.20180507hh](https://doi.org/10.26481/dis.20180507hh)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

Public mental health can be defined as “a public health or population health science approach to mental health and the mental health variations exhibited by populations” (Davies, 2014). Public mental health can be considered as consisting out of three overlapping domains: Mental health promotion, mental illness prevention, and treatment and rehabilitation (Figure 3). At the core of all these domains of public mental health is the measurement and assessment of mental health. Measuring mental health is, however, not a walk in the park. The challenges associated with the measurement of mental health are sometimes evident but are often concealed. They are important nevertheless as conclusions within public mental health, both for individuals but also for policies on different levels, are often (partly) based on the measurement of mental health. At first these challenges may appear small and unimportant, but they can have far-reaching consequences and should therefore be considered the ‘latent defects’ of public mental health. That is, if the public mental health (research) is based on measurements with an unknown blind spot this can have far-reaching implications. The main goal of this thesis was therefore:

- To identify potential latent defects and evaluate their impact within the public mental health

In this summary the rationale, main methods and results of this thesis will be examined and shortly reviewed.

In chapter 1 the conceptualisation of public mental health and its measurement are first introduced. This introduction focusses on four focal points: (1) Mental health moves along a continuum as opposed to a dichotomous representation (healthy/sick). Within a dichotomous representation, or other forms of categorisation, the underlying concept is still a continuum. (2) This underlying continuum is a latent construct. That is, public mental health has no directly observable (manifest) representation. As such public mental health should be inferred from indicators. Measurement instruments use such indicators to grasp the underlying construct. Assessing the quality of such instruments is not straightforward. Quality of a mea-

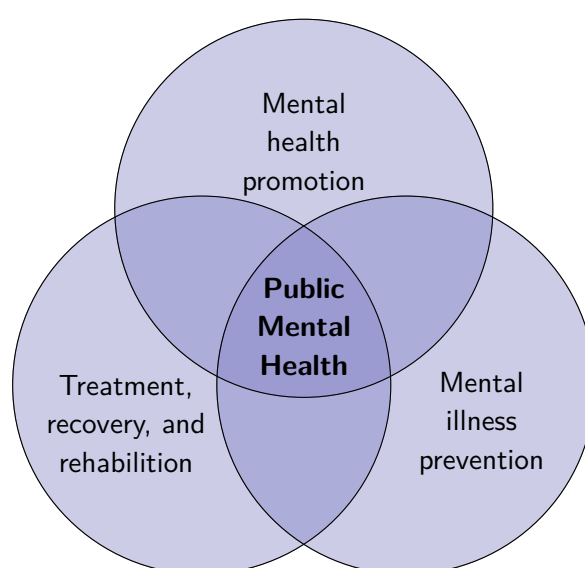


Figure 3: Decomposition of public mental health into its core dimensions, from Davies (2014).

surement instrument has so far mostly been tested on an exploratory and post-hoc basis. This approach has limitations which leaves the application of an instrument vulnerable for possible latent defects. (3) Latent defects are “actors that can result in the violation of (important) quality criteria of an instrument measuring public mental health, without it being (directly) apparent.” The different aspects of this definition were further discussed in the introduction of the thesis. (4) β -Psychometrics are used to assess the quality of a public mental health instrument with an a-priori theoretical framework. Using this framework it is possible to use confirmative statistical testing. With this approach it is, therefore, possible to test specific hypotheses that could rule out (or confirm) the existence of a potential latent defect.

Chapter 2 includes the first empirical study of the thesis. Within this chapter it is examined if the administration of an instrument in different settings could alter the influence of the social desirability bias. Such a difference, in turn, could potentially invalidate the interchangeability of results between the settings. The ‘latent defect’ addressed, in this chapter was the potential misuse of scores (of an instrument) between the different settings. For this chapter the Strengths and Difficulties Questionnaire (SDQ) was used. This is a screening instrument for psychosocial problems in children and adolescents. The compared administration settings were the individual and the collective setting.

The individual setting can be viewed as the clinical application of an instrument (e.g. preventive child healthcare). In this setting an instrument is especially suitable for screening purposes (Vogels, Crone, Hoekstra, & Reijneveld, 2009; van Widenfelt, Goedhart, Treffers, & Goodman, 2003; Goodman, 2001). For this setting data from the preventive child health (PCH) care was used. 6,594 available questionnaires, administered during the screening assessment in the second year of secondary school, were included. In contrast, within the collective setting the information of an instrument is aggregated across a population (van Roy, Groholt, Heyerdahl, & Clench-Aas, 2006; Becker et al., 2006; Ravens-Sieberer, Erhart, Gosch, & Wille, 2008). For this setting information from the National Youth Monitor was used (Ministry of Health, Welfare and Sport & Statistics Netherlands, 2014). This monitor includes the SDQ as an indicator of psychosocial well-being. This resulted in 4,613 available questionnaires. As both samples were retrieved from the same population it was hypothesised that any significant difference in scores between the samples would be a result of the differences in administration between the settings.

Confirmatory factor analysis (CFA) showed that there was no difference in the underlying construct between the two samples. Children in the individual setting, however, scored lower on total difficulties and the psychosocial problems subscales compared to those in the collective setting. To further ‘quantify’ these results the validity of the cross-use between the settings was examined. The cross-use examined was the implementation of cut-off points within the individual setting which are derived from the collective settings. This procedure is comparable with the current practice of implementing cut-off points of the SDQ within the individual setting. Using the 90th percentiles of the collective setting in the individual setting therefore resulted in a small number of cases, 2 to 3%, while ~10% was expected – if there would be no impact of the setting.

These results shows that the SDQ has the same connotation across the individual and collective setting. Nevertheless, the difference in mean scores shows that there is a difference in the ‘trigger-happiness’ within children to tick certain boxes between the setting. This observed structural difference undermines the validity of the cross-use of absolute SDQ-scores between these settings. Applying cut-off scores from the collective setting in the individual setting

could, therefore, result in invalid conclusions and potential misuse of the instrument. To correctly apply cut-off scores these should be retrieved from the applied setting.

The following three chapters (3-5) focus on mental public health in the working population. In these three chapters the same population was used. It is employees who participate in the Maastricht Cohort Study (MCS). It should be noted, however, that the different chapters use (slightly) different subsamples within and between chapters. The MCS was set up in May 1998 and included 12,140 employees from 45 different companies. At this baseline measurement, all included employees were aged between 18 and 65 (Kant et al., 2003; Mohren, Jansen, van Amelsvoort, & Kant, 2007). The baseline (T0) cohort consists of 8,840 (73%) men and 3,255 (27%) women from a working population. For more details regarding the MCS see Mohren et al. (2007). These chapters, as such, focus on the work & health domain of public mental health.

Besides a common sample, these chapters also shared common ground in the public mental health outcomes that were assessed: need for recovery (NFR) and (prolonged) fatigue. NFR is an outcome which captures the need for and extent of recuperation after work. The normal recuperation cycle is ideally completed after the working day or at least before the next working day (van Veldhoven, 2008). Prolonged fatigue is the presence of fatigue, including an overwhelming sense of tiredness, for a more extended period (Bultmann et al., 2000; Kalkman, Zwartz, Schillings, van Engelen, & Bleijenberg, 2008). Both aspects are important public mental health outcomes vital for facilitating or enabling preventive strategies targeting at prevention or limiting further adverse (health) outcomes (e.g., de Croon, Sluiter, & Frings-Dresen, 2003; van Amelsvoort, Kant, Bultmann, & Swaen, 2003; de Raeve, Kant, Jansen, Vasse, & van den Brandt, 2009; van Amelsvoort, Kant, Beurskens, Schroer, & Swaen, 2002; Silva-Costa, Griep, Fischer, & Rotenberg, 2012). NFR was measured with the NFR scale (van Veldhoven, 2003) and prolonged fatigue with the Checklist Individual Strength (Vercoulen et al., 1994; Beurskens et al., 2000). The NFR scale and the CIS have already showed to be instruments with good psychometric properties (e.g., van Veldhoven, 2003; Bultmann et al., 2000; Beurskens et al., 2000; Sluiter, Frings-Dresen, van der Beek, & Meijman, 2001; van Amelsvoort et al., 2003). While these studies show that both measures seem valid measurement instruments many underlying assumptions (latent defect) are not readily addressed.

The first chapter of this 'trilogy' is chapter 3. This chapter examined if the NFR scale and the CIS are measured invariantly across subgroups. Absence of measurement invariance is a typical latent defect as this makes comparisons between groups invalid. Absence of measurement invariance between groups is, however, difficult to observe with classical psychometric tests (Vandenberg & Lance, 2000; Millsap, 2011). Using CFA it was examined if the factor structure was comparable between different subgroups. These subgroups were based on demographic factors (e.g. gender), personal factors (e.g. perceived health), and work-related factors (e.g. psychological job demands). To do so first the hypothesised underlying factor structure was assessed. While many studies examined the factor structure post-hoc, this was the first study to use CFA to assess the factor structure with an a-priori model. With the inclusion of some small adaptations, the hypothesised 1-factor structure (see van Veldhoven, 2008) of the NFR scale showed good fit. For the CIS the hypothesised 4-factor structure (see Janssen, Kant, Swaen, Jansen, & Schroer, 2003) showed good fit, also with the inclusion of some small adaptations. Comparing these baseline models within the different factors (i.e. demographic, personal, and work-related) showed that there was no indication of measurement variance

between any of the subgroups of any of these factors. Chapter 3, as such, shows that the factor structure hypothesised in the literature can be confirmed on an a-priori basis and that this factor structure is comparable across a myriad of subgroups that are often the focus within public mental health.

The following chapter, chapter 4, extended these results to examine whether the NFR scale and CIS were also invariant over time. This is an important assumption within instruments which often goes untested – hence a perfect illustration of a ‘latent defect’. Such stability of measurement over time is crucial for monitoring purposes of both instruments. If variance over time in measurement exists, scores cannot be meaningfully compared, hence hindering a valid monitoring. The findings of this chapter show, however, that the CIS and NFR scale are measured invariantly over time. This opens the door for monitoring and the assessment of psychometric properties important for this purpose. The smallest detectable change for both instruments indicates which change, over time, can be deemed statistically significant within an individual. These scores showed that especially within the NFR scale it was difficult to differentiate within an employee who has elevated NFR scores. Reliability was, however, very good for both scales. It was shown, however, that this was (partly) due to the floor effect within the scores of the NFR scale (also over time). It is, therefore, suggested that, ideally, both instruments are placed within a broader range of instruments to effectively monitor the impact of the psychosocial work environment.

The last chapter of this trilogy, chapter 5, delves deeper into this over time stability of the CIS. To correctly understand and grasp this stability it is important to breakdown the state of fatigue into a time-varying (occasion) and -invariant (trait) component. This also applies for its relation with other adverse health outcomes such as sickness absence. Using the trait-state-occasion (TSO) model (Prenoveau et al., 2011; Prenoveau, 2016) this chapter showed that the trait component of fatigue explained 71% of the state of fatigue. The occasion specific component in contrast only accounted for 28%. The remaining 1% was explained by the previous occasion component. These results show that fatigue, or at least the CIS, has a high trait component. An underlying assumption of fatigue seems, however, to be that it is quite reactive to changes in the environment of a person. This study shows that within a population of employees this assumption does not hold. If not correctly identified, such an assumption could lead to a misjudgement of the results and related conclusions would be hampered (Hamaker, Kuiper, & Grasman, 2015).

Besides this decomposition to study the state of fatigue, as measured by the CIS this chapter also examined the relation between the decomposed components and sickness absence. Again, this relation – often believed to be based on changes in fatigue – could be associated with sickness absence. The significance as predictor for future sickness absence of both the occasion and the trait component of fatigue was therefore examined. It was found that the trait component, uniquely, predicted future sickness absence within a (2-year) time interval within the MCS. This was found for all three different (dichotomous) operationalisations of sickness absence that were taken into account: Spell (consecutive sick days; cut-off ≥ 21 days), duration (total number of sick days; cut-off ≥ 42 days), and frequency (number of spells; cut-off ≥ 3 times). As such it can be concluded that fatigue has a considerable time-invariant component. As this component is also related with other adverse health outcomes, preventive measures and treatment should take this breakdown into account.

Chapter 6 includes a simulation study that was designed to validate a newly proposed informative fit index within Bayesian CFA. While all previous chapters used a frequentist framework,

Bayesian CFA models can be advantageous within specific situations over frequentist CFA models. If a multitude of subgroups (e.g. >40) are for example examined within a frequentist framework, it is very difficult to 'detect' sources of variance between the subgroups. Bayesian CFA has well suited methods to deal with such issues, also within large samples (van de Schoot, Kaplan, et al., 2013). Within large sample Bayesian CFA models, however, a valid fit index that provides an objective assessment of model fit is lacking. Based on the root mean square error of approximation (*RMSEA*) from the frequentist framework the Bayesian *RMSEA* (*BRMSEA*) was therefore proposed. This study shows that this *BRMSEA* performs very well within large samples successfully accepting models with no or a minor misspecification and successfully rejecting models with a major source of misspecification. The added value of the *BRMSEA* was demonstrated within an empirical illustration using the MCS. This illustration examined the factor structure of a subscale of the Job Content Questionnaire (Karasek, 1985). It was shown that the classical fit index becomes uninformative within large samples whilst the *BRMSEA* successfully indicated that the proposed factor structure of the Bayesian CFA model resulted in a good fitting model. This could, for example, open the way for examining structural equivalence of this subscale between a multitude of subgroups (e.g. job-title) using the approximate invariance method within Bayesian CFA (van de Schoot, Kluytmans, et al., 2013). In conclusion it can be stated that the *BRMSEA* is well suited to evaluate model fit in large sample BSEM models by taking sample size and model complexity into account.

Within chapter 7, the general discussion, three important aspects of β -psychometrics are highlighted to deal properly with 'latent defects' regarding the measurement of public mental health. The first aspect concerns the identification of latent defects. Within this section it is illustrated that an active modelling approach is an absolute necessity to successfully identify latent defects within public mental health. The second aspect covers awareness of latent defects. Without this awareness latent defects will always prevail and identification will be a continuous up-hill battle. The discussion covers specific aspects that could increase this awareness (e.g., education and critical review of scholarly work). The third, and final aspect, covers the treatment and preventive strategies regarding potential latent defects. Besides these aspects some conceptual considerations of β -psychometrics in general and some specific for this thesis were addressed. These considerations showed that other alternatives could provide a different perspective but would not greatly alter the results and conclusion of this thesis. It was furthermore stressed that a variety of instruments and populations were included. While all chapters have their profound merit within this thesis and individually, some limitations were also addressed. The development within the field of β -psychometrics regarding the implementation of Bayesian statistics was also considered. Fit indices within this field are however of limited value if large samples – which are typical within public mental health – are used. The newly designed *BRMSEA* of chapter 6 was therefore discussed in combination with the potential implication of this new fit index.

This thesis shows, as such, that 'latent defects' can be successfully identified using β -psychometrics. Using a range of advanced tools this thesis showed that awareness for latent defects is crucial, as measurement is present in almost all aspects within public mental health being it in research, clinical practice, or policy making. As β -psychometrics is based on a-priori model, a critical assessment of the underlying assumptions of an instrument – and its concepts ought to measure – is tantamount. If such an understanding is present 'latent defects' can be potentially identified as shown in this thesis. If latent defects prevail, however, they will

be lurking throughout the whole public mental health hindering valid judgment and decision making.

References

- Becker, A., Steinhausen, H. C., Baldursson, G., Dalsgaard, S., Lorenzo, M. J., Ralston, S. J., ... Rothenberger, A. (2006). Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *European Child and Adolescent Psychiatry, 15 Suppl 1*, i56–i62. doi:10.1007/s00787-006-1008-7
- Beurskens, A. J. H. M., Bultmann, U., Kant, IJ., Vercoulen, J. H. M. M., Bleijenberg, G., & Swaen, G. M. H. (2000). Fatigue among working people: Validity of a questionnaire measure. *Occupational and Environmental Medicine, 57*(5), 353–357. doi:10.1136/oem.57.5.353
- Bultmann, U., de Vries, M., Beurskens, A. J. H. M., Bleijenberg, G., Vercoulen, J. H. M. M., & Kant, IJ. (2000). Measurement of prolonged fatigue in the working population: Determination of a cutoff point for the Checklist Individual Strength. *Journal of Occupational Health Psychology, 5*(4), 411–416. doi:10.1037/1076-8998.5.4.411
- Davies, S. C. (2014). *Annual report of the chief medical officer 2013, public mental health priorities: Investing in the evidence*. London: Department of Health.
- de Croon, E. M., Sluiter, J. K., & Frings-Dresen, M. H. (2003). Need for recovery after work predicts sickness absence. *Journal of Psychosomatic Research, 55*(4), 331–339. doi:10.1016/s0022-3999(02)00630-x
- de Raeve, L., Kant, IJ., Jansen, N. W. H., Vasse, R. M., & van den Brandt, P. A. (2009). Changes in mental health as a predictor of changes in working time arrangements and occupational mobility: Results from a prospective cohort study. *Journal of Psychosomatic Research, 66*(2), 137–145. doi:10.1016/j.jpsychores.2008.05.007
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(11), 1337–45. doi:10.1097/00004583-200111000-00015
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. doi:10.1037/a0038889
- Janssen, N., Kant, IJ., Swaen, G. M. H., Janssen, P. P. M., & Schroer, C. A. P. (2003). Fatigue as a predictor of sickness absence: Results from the Maastricht Cohort Study on fatigue at work. *Occupational and Environmental Medicine, 60*, 71i–76. doi:10.1136/oem.60.suppl_1.i71
- Kalkman, J. S., Zwartz, M. J., Schillings, M. L., van Engelen, B. G. M., & Bleijenberg, G. (2008). Different types of fatigue in patients with facioscapulohumeral dystrophy, myotonic dystrophy and HMSN-I: Experienced fatigue and physiological fatigue. *Neurological Sciences, 29*(S2), 238–240. doi:10.1007/s10072-008-0949-7
- Kant, IJ., Bultmann, U., Schroer, K., Beurskens, A., van Amelsvoort, L. G. P. M., & Swaen, G. (2003). An epidemiological approach to study fatigue in the working population: The Maastricht Cohort Study. *Occupational and Environmental Medicine, 60*, 32i–39. doi:10.1136/oem.60.suppl_1.i32
- Karasek, R. A. (1985). *Job Content Questionnaire and user's guide*. University of Massachusetts, Department of Work Environment. Lowell.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

- Ministry of Health, Welfare and Sport & Statistics Netherlands. (2014). Youth monitor. Retrieved from <http://jeugdmonitor.cbs.nl/en-gb/information/>
- Mohren, D. C. L., Jansen, N. W. H., van Amelsvoort, L. P. G. M., & Kant, IJ. (2007). *An epidemiological approach of fatigue at work: Experiences from the Maastricht Cohort Study*. Programma Epidemiologie van Arbeid en Gezondheid Maastricht University, Maastricht.
- Prenoveau, J. M. (2016). Specifying and interpreting latent state-trait models with autoregression: An illustration. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5), 731–749. doi:10.1080/10705511.2016.1186550
- Prenoveau, J. M., Craske, M. G., Zinbarg, R. E., Mineka, S., Rose, R. D., & Griffith, J. W. (2011). Are anxiety and depression just as stable as personality during late adolescence? Results from a three-year longitudinal latent variable study. *Journal of Abnormal Psychology*, 120(4), 832–843. doi:10.1037/a0023939
- Ravens-Sieberer, U., Erhart, M., Gosch, A., & Wille, N. (2008). Mental health of children and adolescents in 12 European countries – results from the European KIDSCREEN study. *Clinical Psychology & Psychotherapy*, 15(3), 154–63. doi:10.1002/cpp.574
- Silva-Costa, A., Griep, R. H., Fischer, F. M., & Rotenberg, L. (2012). Need for recovery from work and sleep-related complaints among nursing professionals. *Work*, 41(S1), 3726–3731. doi:10.3233/WOR-2012-0086-3726
- Sluiter, J. K., Frings-Dresen, M. H., van der Beek, A. J., & Meijman, T. F. (2001). The relation between work-induced neuroendocrine reactivity and recovery, subjective need for recovery, and health status. *Journal of Psychosomatic Research*, 50(1), 29–37. doi:10.1016/s0022-3999(00)00213-0
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2013). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842–860. doi:10.1111/cdev.12169
- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthen, B. O. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00770
- van Amelsvoort, L. G. P. M., Kant, IJ., Beurskens, A. J. H. M., Schroer, C. A. P., & Swaen, G. H. M. (2002). Fatigue as a predictor of work disability. *Occupational and Environmental Medicine*, 59(10), 712–713. doi:10.1136/oem.59.10.712
- van Amelsvoort, L. G. P. M., Kant, IJ., Bultmann, U., & Swaen, G. M. H. (2003). Need for recovery after work and the subsequent risk of cardiovascular disease in a working population. *Occupational and Environmental Medicine*, 60(90001), i83–i87. doi:10.1136/oem.60.suppl_1.i83
- van Roy, B., Groholt, B., Heyerdahl, S., & Clench-Aas, J. (2006). Self-reported strengths and difficulties in a large Norwegian population 10-19 years: Age and gender specific results of the extended SDQ-questionnaire. *European Child and Adolescent Psychiatry*, 15(4), 189–98. doi:10.1007/s00787-005-0521-4
- van Veldhoven, M. (2003). Measurement quality and validity of the Need for Recovery Scale. *Occupational and Environmental Medicine*, 60, 3i–9. doi:10.1136/oem.60.suppl_1.i3
- van Veldhoven, M. (2008). Need for recovery after work: An overview of construct, measurement and research. In J. Houdmont & S. Leka (Eds.), *Occupational health psychology. european perspectives on research, education and practice*. (pp. 1–25). Nottingham: Nottingham University Press.

- van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry, 12*(6), 281–9. doi:10.1007/s00787-003-0341-3
- 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. doi:10.1177/109442810031002
- Vercoulen, J. H., Swanink, C. M., Fennis, J. F., Galama, J. M., van der Meer, J. W., & Bleijenberg, G. (1994). Dimensional assessment of chronic fatigue syndrome. *Journal of Psychosomatic Research, 38*(5), 383–392. doi:10.1016/0022-3999(94)90099-x
- Vogels, A. G., Crone, M. R., Hoekstra, F., & Reijneveld, S. A. (2009). Comparing three short questionnaires to detect psychosocial dysfunction among primary school children: A randomized method. *BMC Public Health, 9*, 489. doi:10.1186/1471-2458-9-489

Samenvatting

Publieke mentale gezondheid kan worden gedefinieerd als “een benadering op basis van de publieke gezondheidszorg van de mentale gezondheid en variatie binnen de populatie van de mentale gezondheid (Davies, 2014).” Publieke mentale gezondheid bestaat uit drie overlap-pende domeinen: Aandacht voor mentale gezondheid, preventie van psychische aandoeningen, en behandeling en revalidatie. Startpunt van al deze drie domeinen omvat meestal het meten en beoordelen van mentale gezondheid. Meten van mentale gezondheid is echter niet eenvoudig omdat aandoeningen zich meestal niet in aan- of afwezigheid manifesteren maar in een continuüm van licht tot matig of ernstig. De uitslag van de meting is dus niet dichtoorn maar bestrijkt een breed spectrum. De gevolgen van de metingsuitslag zijn verstrekkend en resulteren op individueel niveau in een behandelingstraject en op populatieniveau in een bepaalde beleidskeuze. Het correct meten van mentale gezondheid is daarom van groot belang. De uitdaging betreft vooral ook het tijdig opsporen van potentiële verborgen gebreken. Als (onderzoek naar) publieke mentale gezondheid (mede) gebaseerd wordt op meetinstrumenten met een onbekende blinde vlek, kan dit een negatieve impact hebben. Het hoofddoel van dit proefschrift was daarom:

- Het identificeren van potentiële verborgen gebreken en de impact hiervan evalueren op de publieke mentale gezondheid

In deze samenvatting worden de motivering, methodes, en resultaten zoals gebruikt in de verschillende hoofdstukken kort uiteengezet en besproken.

In hoofdstuk 1 is het concept publieke mentale gezondheid en het meten ervan voor het eerst geïntroduceerd. De introductie richt zich op vier belangrijke punten: (1) Mentale gezondheid is een continuüm in plaats van een dichotome representatie (ziek/gezond). Zelfs als een dichotome representatie wordt gebruikt, of een andere vorm van categorisatie, is het onderliggende construct nog steeds continu. (2) Dit onderliggende continuüm is latent van aard. Dit wil zeggen dat publieke mentale gezondheid geen direct meetbare (manifeste) fenomenen kent. Daarom moet publieke mentale gezondheid worden gemeten aan de hand van indicatoren. Meetinstrumenten gebruiken dergelijke indicatoren om het onderliggende construct te omvatten. Het bepalen van de kwaliteit van dergelijke instrumenten is niet eenvoudig. Tot dusver wordt de kwaliteit van meetinstrumenten voornamelijk getest door middel van verkennende en ad-hoc analyses. Een dergelijke aanpak kent verschillende beperkingen wat er voor zorgt dat de meetinstrumenten kwetsbaar zijn voor verborgen gebreken. (3) Verborgen gebreken zijn “actoren die resulteren in de schending van (belangrijke) kwaliteitscriteria van een meetinstrument binnen publieke mentale gezondheid, zonder dat dit (direct) duidelijk is.” De verschillende aspecten in deze definitie worden verder bediscussieerd in de introductie van dit proefschrift. (4) β -psychometrie kan worden toegepast om de kwaliteit van instrumenten die publieke mentale gezondheid meten te evalueren door middel van een a-priori theoretisch kader. Dit theoretisch kader maakt het mogelijk om confirmatief te testen. Hierdoor is het mogelijk om specifieke hypothesen te testen op de aan- of afwezigheid van verborgen gebreken.

Hoofdstuk 2 vormt de eerste empirische studie van het proefschrift. In dit hoofdstuk wordt onderzocht of het afnemen van een meetinstrument in verschillende settings, gevolgen heeft voor het effect van sociale wenselijkheid waardoor de uitkomst verandert. Dit zou er voor kunnen zorgen dat de vergelijking tussen deze verschillende settings niet valide is. Het ‘verborgen gebrek’ dat hier wordt onderzocht was dus het potentieel verkeerde gebruik van scores (van een meetinstrument) tussen settings. In dit hoofdstuk is gebruik gemaakt van de Nederlandse versie van de ‘Strengths and Difficulties Questionnaire’ (SDQ; Goodman, 1997). Dit is een

screeningsinstrument voor psychosociale problemen bij kinderen en adolescenten. Hierbij is de afname in een individuele setting vergeleken met de afname in een collectieve setting.

De individuele setting kan worden gezien als de één-op-één toepassing van een instrument (bijv. van de jeugdgezondheidszorg). In deze setting is een instrument vooral geschikt voor screeningsdoeleinden (Vogels, Crone, Hoekstra & Reijneveld, 2009; van Widenfelt, Goedhart, Treffers & Goodman, 2003; Goodman, 2001). Voor beide settings is gebruik gemaakt van data afkomstig van de jeugdgezondheidszorg. Voor de individuele setting zijn 6.594 vragenlijsten geanalyseerd die afgenomen zijn tijdens het periodiek gezondheidsonderzoek tijdens het tweede jaar van de middelbare school. In de collectieve setting wordt de informatie van een meetinstrument juist geaggregeerd binnen een populatie (van Roy, Groholt, Heyerdahl & Clench-Aas, 2006; Becker e.a., 2006; Ravens-Sieberer, Erhart, Gosch & Wille, 2008). Voor de collectieve setting is gebruik gemaakt van de informatie uit de jeugdmonitor die anoniem afgenomen wordt in klas 2 van het voortgezet onderwijs (Ministerie van Volksgezondheid, Welzijn en Sport & Centraal Bureau voor de Statistiek, 2014). Deze monitor bevat onder andere de SDQ als een indicator van psychosociaal welbevinden. Dit resulteert in de analyse van 4.613 vragenlijsten. Aangezien beide steekproeven uit dezelfde populatie voortkomen, is het aannemelijk dat eventuele significante verschillen een gevolg zijn van verschillen in de setting van de afname.

Confirmatieve factor analyses (CFA) laten zien dat er geen significante verschillen zijn in het onderliggende construct (*psychosociale problemen*) vergeleken tussen de twee steekproeven. Kinderen in de individuele setting, scoren desalniettemin lager wat betreft totaal probleemgedrag en de onderliggende subschalen in vergelijking met de kinderen in de collectieve setting. Om deze resultaten verder te kwantificeren is er gekeken naar de vergelijkbaarheid van de scores in de verschillende settings. Dit is onderzocht door het toepassen van afkappunten, gebaseerd op de collectieve setting, binnen de individuele setting. Deze procedure is vergelijkbaar met staand beleid aangaande het gebruik van afkappunten van de SDQ binnen de jeugdgezondheidszorg. Hierbij resulteert het gebruik van het 90^e percentiel uit de collectieve setting binnen de individuele setting in slechts een klein aantal gevallen met een afwijkende score (2 tot 3%). Dit terwijl 10% zou worden verwacht – als de setting geen impact zou hebben op de afname.

Deze resultaten laten zien dat de SDQ dezelfde connotatie heeft in de individuele en collectieve setting. Het verschil in de gemiddelde scores laat echter zien dat er een verschil zit in de bereidwilligheid om een bepaald vakje aan te kruisen bij afname in de verschillende settings. Dit geobserveerde structurele verschil ondermijnt de validiteit van onderling gebruik van scores tussen settings. De toepassing van de afkappunten, afkomstig uit de collectieve setting, binnen de individuele setting resulteert in conclusies die niet valide zijn. Om deze afkappunten correct toe te passen zullen ze daarom ook afzonderlijk binnen iedere setting moeten worden vastgesteld.

De hoofdstukken 3-5 richten zich op de mentale gezondheid van werkenden. De drie studies maakten allen gebruik van dezelfde studiepopulatie, namelijk de deelnemers aan de Maastrichtse Cohort Studie (MCS). Binnen de verschillende hoofdstukken wordt er echter wel gebruik gemaakt van verschillende selecties binnen deze populatie. De MCS is opgezet in mei 1998 en betrof toen 12.140 werknemers van 45 verschillende bedrijven. Gedurende deze baseline meting (T0) waren alle werknemers tussen 18 en 65 jaar oud, waarvan 8.840 man (73%) en 3.255 vrouw (Kant e.a., 2003; Mohren, Jansen, van Amelsvoort & Kant, 2007). Voor meer details aangaande de MCS zie Mohren e.a. (2007).

Naast dezelfde steekproef delen deze hoofdstukken ook de uitkomstmaten die bestudeerd zijn: herstelbehoefte en (langdurige) vermoeidheid. Herstelbehoefte is een uitkomstmaat welke zich vooral richt op de behoefte en duur van herstel na een werkdag. De normale herstelcyclus is idealiter afgerond na het werk maar in ieder geval voor het begin van de volgende werkdag (van Veldhoven, 2008). Langdurige vermoeidheid betreft de aanwezigheid van vermoeidheid, inclusief een overweldigend gevoel van vermoeidheid, voor een langere periode (Bultmann e.a., 2000; Kalkman, Zwartz, Schillings, van Engelen & Bleijenberg, 2008). Beide aspecten zijn belangrijke uitkomstmaten binnen de publieke mentale gezondheid. Daarnaast kunnen deze aspecten een belangrijke rol spelen bij preventieve strategieën die gericht zijn op het voorkomen of beperken van verdere negatieve (gezondheids-) uitkomsten (bijv., de Croon, Sluiter & Frings-Dresen, 2003; van Amelsvoort, Kant, Bultmann & Swaen, 2003; de Raeve, Kant, Jansen, Vasse & van den Brandt, 2009; van Amelsvoort, Kant, Beurskens, Schroer & Swaen, 2002; Silva-Costa, Griep, Fischer & Rotenberg, 2012). Herstelbehoefte wordt gemeten met de herstelbehoefteschaal (van Veldhoven, 2003) en langdurige vermoeidheid met de Checklist Individuele Spankracht (Vercoulen e.a., 1994; Beurskens e.a., 2000). De herstelbehoefteschaal en de CIS zijn beide gevalideerde meetinstrumenten met goede psychometrische eigenschappen (zie bijv., van Veldhoven, 2003; Bultmann e.a., 2000; Beurskens e.a., 2000; Sluiter, Frings-Dresen, van der Beek & Meijman, 2001; van Amelsvoort e.a., 2003). Ook al laten verschillende studies zien dat het valide meetinstrumenten betreft, verschillende onderliggende assumpties – *verborgen gebreken* – zijn (nog) niet adequaat onderzocht.

In hoofdstuk 3, het eerste deel van deze trilogie, wordt onderzocht of de herstelbehoefteschaal en de CIS al dan niet variabel functioneren in verschillende subgroepen. Variabele functioneren van een instrument tussen groepen is een typisch verborgen gebrek aangezien dit resulteert in een niet valide vergelijking tussen groepen. Afwezigheid van meetinvariantie tussen groepen is echter moeilijk vast te stellen middels klassieke psychometrische analyses (Vandenberg & Lance, 2000; Millsap, 2011). Met CFA is daarom onderzocht of de factor structuur vergelijkbaar is tussen verschillende subgroepen. Deze subgroepen zijn gebaseerd op demografische factoren (bijv. geslacht), persoonlijke factoren (bijv. ervaren gezondheid), en werk gerelateerde factoren (bijv. soort arbeidspatroon). Om dit te analyseren is eerst de onderliggende factor structuur geëvalueerd. Ook al hebben vorige studies deze factor structuur al op een post-hoc basis bekeken, dit is de eerste studie die CFA gebruikt om de factor structuur te evalueren op basis van een a-priori model. Met een aantal kleine aanpassingen laat de gehypothetiseerde 1-factor structuur van de herstelbehoefteschaal een goed passend model zien (van Veldhoven, 2008). Ook voor de CIS laat de gehypothetiseerde 4-factor structuur met een aantal kleine aanpassingen een goed passend model zien (Janssen, Kant, Swaen, Janssen & Schroer, 2003). Alle modellen worden vervolgens vergeleken met deze baseline-modellen. Bij deze vergelijking is er geen aanwijzing voor meetvariantie tussen groepen. Hoofdstuk 3, toont daarmee aan dat op een a-priori basis de gehypothetiseerde factor structuur kan worden bevestigd. Daarnaast toont dit hoofdstuk aan dat deze factor structuur vergelijkbaar is voor een groot aantal subgroepen die vaak worden vergeleken binnen de publieke mentale gezondheid.

In het hierop volgende deel, hoofdstuk 4, wordt dieper ingegaan op deze resultaten door te onderzoeken of er ook sprake is van meetinvariantie over tijd. Dit is een belangrijke assumptie bij meetinstrumenten die vaak niet getest wordt – hierdoor is dit een perfecte illustratie van een ‘verborgen gebrek’. Dergelijke invariantie van metingen over tijd is cruciaal voor het adequaat monitoren van bijvoorbeeld herstelbehoefte en langdurige vermoeidheid. Als meet-

variantie namelijk aanwezig is dan kunnen scores niet inhoudelijk met elkaar worden vergeleken waardoor kwalitatieve monitoring onmogelijk is. Hoofdstuk 4 laat zien dat zowel de CIS als de herstelbehoefteschaal geen meetvariantie hebben over tijd. Hierdoor is het dus mogelijk om deze belangrijke concepten te monitoren en evalueren. Daarnaast wordt er ook gekeken naar het kleinste verschil in score dat kan worden onderscheiden van meetfouten. Deze score kan dus worden gezien als een significant verschil binnen een persoon over tijd. Deze analyse laat zien dat het vooral bij de herstelbehoefteschaal moeilijk is om te differentiëren tussen werknemers met een verhoogde herstelbehoefte. De betrouwbaarheid van beide meetinstrumenten blijkt erg goed. Dit komt echter deels door het vloer effect dat optreedt bij de herstelbehoefteschaal. De aanbeveling is daarom ook dat idealiter dergelijke meetinstrumenten worden geïncorporeerd binnen een bredere testbatterij om zodoende de effecten van de psychosociale werkomgeving betrouwbaar te meten

In het laatste deel van deze 'trilogie', hoofdstuk 5, wordt dieper ingegaan op de longitudinale stabiliteit van de CIS. Om dit goed te kunnen duiden is het van belang om de geobserveerde toestand van vermoeidheid te splitsen in een component die varieert over tijd en een component die stabiel is over tijd. Daarnaast is er ook gekeken naar de relatie met andere negatieve gezondheidsuitkomsten zoals ziekteverzuim. Door gebruik te maken van het *Trait-State-Occasion* (TSO) model (Prenoveau e.a., 2011; Prenoveau, 2016) in dit hoofdstuk wordt aangetoond dat de tijd-invariante component van vermoeidheid 71% van de geobserveerde vermoeidheid verklaart. De tijd-variante component daarentegen verklaart slechts 28%. De overgebleven 1% wordt verklaard door het vorige meetmoment. Deze resultaten laten zien dat vermoeidheid, of toch in ieder geval de CIS, een hoge mate van stabiliteit over tijd kent. Een onderliggende aanname bij vermoeidheid lijkt echter te zijn dat het redelijk reactief is op veranderingen in de omgeving van een persoon. Dit hoofdstuk toont echter aan dat binnen een steekproef van werknemers deze assumptie niet wordt onderschreven. Als deze assumptie echter zonder kritische toetsing als startpunt wordt gebruikt, kan dit leiden tot resultaten en conclusies die niet valide zijn (Hamaker, Kuiper & Grasman, 2015).

Naast deze decompositie om de geobserveerde toestand van vermoeidheid te bestuderen, gemeten met de CIS, wordt er in dit hoofdstuk ook de relatie tussen deze verschillende componenten en ziekteverzuim geanalyseerd. Er is daarom gekeken naar de tijd-invariante en tijd-variante component in relatie met ziekteverzuim. Uit de analyses blijkt dat alleen de tijd-invariante component ziekteverzuim voorspelt binnen een 2-jaar durend tijdsinterval. Dit is zo voor drie verschillende dichotome operationalisaties van ziekteverzuim: een langdurige verzuimperiode (opeenvolgende ziekte-dagen, ≥ 21 dagen), lange totale duur (totaal aantal ziekte-dagen, ≥ 42 dagen) en een hoge frequentie (aantal periodes; \geq drie periodes). De tijd-invariante component van vermoeidheid heeft dus de grootste voorspellende waarde met betrekking tot negatieve gezondheidseffecten. Aangezien ook al geconcludeerd is dat deze component een groot deel van de variantie van vermoeidheid verklaart, is het dus belangrijk dat preventieve maatregelen en behandelingen deze decompositie in ogenschouw nemen.

Hoofdstuk 6 beschrijft een simulatie studie die was opgezet om een nieuw ontworpen passings-index binnen Bayesiaanse CFA te valideren. Daar waar alle vorige hoofdstukken een frequentistisch kader gebruikten, kan Bayesiaanse CFA verschillende voordelen bieden in bepaalde situaties in vergelijking met frequentistische modellen. Wanneer er bijvoorbeeld een veelvoud aan subgroepen (bijv. > 40) is, kan het binnen een frequentistisch kader erg moeilijk zijn om de bron van eventuele verschillen te 'ontdekken'. Bayesiaanse CFA heeft goede methoden om met dergelijke problemen om te gaan, ook als de steekproef groot is (van de Schoot, Kaplan

e.a., 2013). Binnen dergelijke Bayesiaanse CFA-modellen, met een grote steekproef, ontbreekt het echter aan een bruikbare passings-index die een objectieve beoordeling biedt voor de kwaliteit van een model. Op basis van de root mean square error of approximation (*RMSEA*) uit het frequentistische kader is daarom de Bayesiaanse *RMSEA* (*BRMSEA*) voorgesteld. Dit hoofdstuk toont aan dat deze *BRMSEA* zeer goed presteert in grote steekproeven en succesvol modellen accepteert met geen of een kleine misspecificatie en modellen afkeurt met significante misspecificatie(s). De toegevoegde waarde van de *BRMSEA* wordt verder aangetoond in een empirische illustratie met behulp van de MCS. Deze illustratie onderzoekt de factorstructuur van een subschaal van de Job Content Questionnaire (Karasek, 1985). Hier is aangetoond dat de klassieke passings-index (de *ppp*) niet-informatief is binnen grote steekproeven, terwijl de *BRMSEA* succesvol detecteert dat de voorgestelde factorstructuur een goed passend model is. De *BRMSEA* maakt het dus mogelijk om de aanwezigheid van meetinvariantie te onderzoeken bij meetinstrumenten tussen een groot aantal subgroepen (van de Schoot, Kluytmans e.a., 2013). Concluderend kan worden vastgesteld dat de *BRMSEA* goed geschikt is om in BSEM modellen met een grote steekproef de kwaliteit van een model te evalueren door rekening te houden met deze steekproefomvang en de complexiteit van het (factor-)model.

In hoofdstuk 7, de algemene discussie, worden drie belangrijke aspecten benadrukt om goed om te gaan met 'verborgen gebreken' in relatie tot het meten van de publieke mentale gezondheid. Het eerste aspect betreft de identificatie/opsoring van verborgen gebreken. Hierbij is geïllustreerd dat een actieve manier van modelleren een absolute noodzaak is om succesvol verborgen gebreken binnen de publieke mentale gezondheid te identificeren. Het tweede aspect betreft het vergroten van het bewustzijn voor dergelijke verborgen gebreken. Zonder dit bewustzijn zullen verborgen gebreken altijd de overhand hebben. De discussie gaat ook in op specifieke methoden om dit bewustzijn te vergroten (bijv. onderwijs en kritische beoordeling van wetenschappelijk werk). Het derde en laatste aspect betreft de aanpak en preventie van potentiële verborgen gebreken aangaande het meten van publieke mentale gezondheid. Naast deze aspecten worden enkele conceptuele eigenschappen van β -psychometrie behandeld. Hierbij wordt zowel gekeken naar meer generieke eigenschappen als eigenschappen die specifiek voor dit proefschrift van belang zijn. Hierbij wordt aangetoond dat andere methodes een alternatief perspectief kunnen bieden op de conclusies en bevindingen van dit proefschrift. Verder wordt benadrukt dat een verscheidenheid aan instrumenten en populaties is bestudeerd. Ook komen enkele beperkingen van de verschillende hoofdstukken aan bod. De ontwikkeling op het gebied van β -psychometrie met betrekking tot de implementatie van Bayesiaanse statistiek is ook aangestipt. Momenteel zijn passings-indexen hierbinnen echter nog niet erg bruikbaar bij grote steekproeven – welke typisch zijn voor de publieke mentale gezondheid. De *BRMSEA* in hoofdstuk 6 kan hier verandering in brengen.

Dit proefschrift toont aan dat 'verborgen gebreken' met succes kunnen worden geïdentificeerd door middel van β -psychometrie. Met behulp van een variëteit aan geavanceerde analyses maken de verschillende studies die zijn beschreven in dit proefschrift duidelijk dat een toenemend bewustzijn voor verborgen gebreken cruciaal is. Metingen liggen immers aan de basis van bijna alle aspecten binnen de publieke mentale gezondheid, betreffende onderzoek, klinische praktijk en beleidsvorming. Aangezien β -psychometrie gebaseerd is op a-priori modellen is het van groot belang dat de onderliggende aannames van een meetinstrument worden geëxpliciteerd alvorens het mogelijk is de kwaliteit van een meetinstrument te beoordelen. Als een dergelijk begripsniveau met betrekking tot een meetinstrument aanwezig is, kunnen 'ver-

borgen gebreken' worden geïdentificeerd. Dit is belangrijk omdat verborgen gebreken anders een negatieve uitwerking hebben binnen de gehele publieke mentale gezondheid, waardoor een valide oordeel en besluitvorming worden belemmerd.

Literatuur

- Becker, A., Steinhausen, H. C., Baldursson, G., Dalsgaard, S., Lorenzo, M. J., Ralston, S. J., ... Rothenberger, A. (2006). Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *European Child and Adolescent Psychiatry, 15 Suppl 1*, i56–i62. doi:10.1007/s00787-006-1008-7
- Beurskens, A. J. H. M., Bultmann, U., Kant, IJ., Vercoulen, J. H. M. M., Bleijenberg, G. & Swaen, G. M. H. (2000). Fatigue among working people: Validity of a questionnaire measure. *Occupational and Environmental Medicine, 57(5)*, 353–357. doi:10.1136/oem.57.5.353
- Bultmann, U., de Vries, M., Beurskens, A. J. H. M., Bleijenberg, G., Vercoulen, J. H. M. M. & Kant, IJ. (2000). Measurement of prolonged fatigue in the working population: Determination of a cutoff point for the Checklist Individual Strength. *Journal of Occupational Health Psychology, 5(4)*, 411–416. doi:10.1037/1076-8998.5.4.411
- Davies, S. C. (2014). *Annual report of the chief medical officer 2013, public mental health priorities: Investing in the evidence*. London: Department of Health.
- de Croon, E. M., Sluiter, J. K. & Frings-Dresen, M. H. (2003). Need for recovery after work predicts sickness absence. *Journal of Psychosomatic Research, 55(4)*, 331–339. doi:10.1016/s0022-3999(02)00630-x
- de Raeve, L., Kant, IJ., Jansen, N. W. H., Vasse, R. M. & van den Brandt, P. A. (2009). Changes in mental health as a predictor of changes in working time arrangements and occupational mobility: Results from a prospective cohort study. *Journal of Psychosomatic Research, 66(2)*, 137–145. doi:10.1016/j.jpsychores.2008.05.007
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38(5)*, 581–6. doi:10.1111/j.1469-7610.1997.tb01545.x
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40(11)*, 1337–45. doi:10.1097/00004583-200111000-00015
- Hamaker, E. L., Kuiper, R. M. & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20(1)*, 102–116. doi:10.1037/a0038889
- Janssen, N., Kant, IJ., Swaen, G. M. H., Janssen, P. P. M. & Schroer, C. A. P. (2003). Fatigue as a predictor of sickness absence: Results from the Maastricht Cohort Study on fatigue at work. *Occupational and Environmental Medicine, 60*, 71i–76. doi:10.1136/oem.60.suppl_1.i71
- Kalkman, J. S., Zwarts, M. J., Schillings, M. L., van Engelen, B. G. M. & Bleijenberg, G. (2008). Different types of fatigue in patients with facioscapulohumeral dystrophy, myotonic dystrophy and HMSN-I: Experienced fatigue and physiological fatigue. *Neurological Sciences, 29(S2)*, 238–240. doi:10.1007/s10072-008-0949-7
- Kant, IJ., Bultmann, U., Schroer, K., Beurskens, A., van Amelsvoort, L. G. P. M. & Swaen, G. (2003). An epidemiological approach to study fatigue in the working population:

- The Maastricht Cohort Study. *Occupational and Environmental Medicine*, 60, 32i–39. doi:10.1136/oem.60.suppl_1.i32
- Karasek, R. A. (1985). *Job Content Questionnaire and user's guide*. University of Massachusetts, Department of Work Environment. Lowell.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.
- Ministerie van Volksgezondheid, Welzijn en Sport & Centraal Bureau voor de Statistiek. (2014). Jeugdmonitor.
- Mohren, D. C. L., Jansen, N. W. H., van Amelsvoort, L. P. G. M. & Kant, IJ. (2007). *An epidemiological approach of fatigue at work: Experiences from the Maastricht Cohort Study*. Programma Epidemiologie van Arbeid en Gezondheid Maastricht University, Maastricht.
- Prenoveau, J. M. (2016). Specifying and interpreting latent state-trait models with auto-regression: An illustration. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5), 731–749. doi:10.1080/10705511.2016.1186550
- Prenoveau, J. M., Craske, M. G., Zinbarg, R. E., Mineka, S., Rose, R. D. & Griffith, J. W. (2011). Are anxiety and depression just as stable as personality during late adolescence? Results from a three-year longitudinal latent variable study. *Journal of Abnormal Psychology*, 120(4), 832–843. doi:10.1037/a0023939
- Ravens-Sieberer, U., Erhart, M., Gosch, A. & Wille, N. (2008). Mental health of children and adolescents in 12 European countries – results from the European KIDSCREEN study. *Clinical Psychology & Psychotherapy*, 15(3), 154–63. doi:10.1002/cpp.574
- Silva-Costa, A., Griep, R. H., Fischer, F. M. & Rotenberg, L. (2012). Need for recovery from work and sleep-related complaints among nursing professionals. *Work*, 41(S1), 3726–3731. doi:10.3233/WOR-2012-0086-3726
- Sluiter, J. K., Frings-Dresen, M. H., van der Beek, A. J. & Meijman, T. F. (2001). The relation between work-induced neuroendocrine reactivity and recovery, subjective need for recovery, and health status. *Journal of Psychosomatic Research*, 50(1), 29–37. doi:10.1016/s0022-3999(00)00213-0
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J. & van Aken, M. A. G. (2013). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842–860. doi:10.1111/cdev.12169
- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J. & Muthen, B. O. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00770
- van Amelsvoort, L. G. P. M., Kant, IJ., Beurskens, A. J. H. M., Schroer, C. A. P. & Swaen, G. H. M. (2002). Fatigue as a predictor of work disability. *Occupational and Environmental Medicine*, 59(10), 712–713. doi:10.1136/oem.59.10.712
- van Amelsvoort, L. G. P. M., Kant, IJ., Bultmann, U. & Swaen, G. M. H. (2003). Need for recovery after work and the subsequent risk of cardiovascular disease in a working population. *Occupational and Environmental Medicine*, 60(90001), i83–i87. doi:10.1136/oem.60.suppl_1.i83
- van Roy, B., Groholt, B., Heyerdahl, S. & Clench-Aas, J. (2006). Self-reported strengths and difficulties in a large Norwegian population 10-19 years: Age and gender specific results of the extended SDQ-questionnaire. *European Child and Adolescent Psychiatry*, 15(4), 189–98. doi:10.1007/s00787-005-0521-4
- van Veldhoven, M. (2003). Measurement quality and validity of the Need for Recovery Scale. *Occupational and Environmental Medicine*, 60, 3i–9. doi:10.1136/oem.60.suppl_1.i3

- van Veldhoven, M. (2008). Need for recovery after work: An overview of construct, measurement and research. In J. Houdmont & S. Leka (Red.), *Occupational health psychology. european perspectives on research, education and practice*. (P. 1–25). Nottingham: Nottingham University Press.
- van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry*, 12(6), 281–9. doi:10.1007/s00787-003-0341-3
- 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. doi:10.1177/109442810031002
- Vercoulen, J. H., Swanink, C. M., Fennis, J. F., Galama, J. M., van der Meer, J. W. & Bleijenberg, G. (1994). Dimensional assessment of chronic fatigue syndrome. *Journal of Psychosomatic Research*, 38(5), 383–392. doi:10.1016/0022-3999(94)90099-x
- Vogels, A. G., Crone, M. R., Hoekstra, F. & Reijneveld, S. A. (2009). Comparing three short questionnaires to detect psychosocial dysfunction among primary school children: A randomized method. *BMC Public Health*, 9, 489. doi:10.1186/1471-2458-9-489