

# All that looks grave is not grievous. Not all those who wince are in pain

Citation for published version (APA):

van Impelen, A. (2018). *All that looks grave is not grievous. Not all those who wince are in pain: studies in furtherance of validity assessment*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20181025ai>

## Document status and date:

Published: 01/01/2018

## DOI:

[10.26481/dis.20181025ai](https://doi.org/10.26481/dis.20181025ai)

## 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](http://www.umlib.nl/taverne-license)

## Take down policy

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

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

## Summary



The research presented in this thesis aims to advance the evaluation of assessment validity. This goal was pursued through empirical investigations of indicators (Chapters 2–3) and tests (Chapters 4–6) that are considered to be measures of assessment validity. By way of introduction, Chapter 1 touches upon the historical context of validity testing and brings up several common myths and misconceptions. Chapter 7 discusses the main findings of the studies presented throughout this thesis and offers considerations of methodological issues, clinical implications, and directions for future research. To enhance the utility of this thesis and advance the practice of validity assessment, Chapter 7 also includes a primer on a quantitative method for interpreting and combining validity test data; the likelihood ratio.

## Chapter 1

Chapter 1 introduces the validity triumvirate that determines the accuracy of diagnoses; test validity, assessment validity, and judgment validity. Test validity refers to the alignment between what a test measures and what it is designed and employed to measure. Assessment validity concerns the accuracy and openness with which patients present their symptoms. Judgment validity denotes the extent to which diagnostic conclusions follow logically from clinical data. Chapter 1 also conveys that diagnostic decision-making can be accomplished through two methodological approaches that are fundamentally different; the clinical method and the actuarial method: A decision-making process is clinical if it depends on human cognition; it is actuarial if it solely rests upon empirically derived decision rules.

Before divulging the present state of play in validity assessment, Chapter 1 serves up historical highlights relevant to some of the myths and misconceptions that vex psychological and neuropsychological assessments still. Foremost amongst these myths and misconceptions are the ungrounded belief in the potency of clinical judgment in validity assessment; the incorrect notion that feigned symptom presentations are rare; the false idea that feigned symptomatology must be underlain by authentic symptomatology; and the faulty perception that antisocial features are associated with feigned symptomatology.

Finally, Chapter 1 describes how the modern-day approach to validity assessment originated from attempts to use two-alternative forced-choice tests to identify functional neurological deficits. Nowadays, the evaluation of assessment validity is realized by means of “performance validity test” (PVTs) and “symptom validity tests” (SVTs): PVTs resemble cognitive tests and are used to appraise the adequacy of effort that examinees give during cognitive testing. SVTs are similar to symptom inventories or interviews, and serve to estimate the accuracy with which patients present their symptoms. An important constraint of validity test is that they cannot identify *why* assessment validity is insufficient; they can only indicate *whether* it is insufficient.

## Chapter 2

Chapter 2 covers two studies, both of which were designed to examine the idea, as articulated in the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.), that antisocial

features and behavior are predictive of feigned symptom presentations. In the first study, we administered an SVT and a measure of symptom minimization to 57 forensic psychiatric inpatients, and rated the behavior of participants by coding a behavioral observation scale based on the daily logs in patient records. In the second study, we presented 45 inmates of a youth prison with an SVT, and with self-report scales for psychopathy and criminal blame attribution. We inferred antisocial behavior of prisoners from the length of their sentence, from whether or not they were officially considered to be intensive offenders, and from the number of disciplinary actions they incurred while incarcerated.

The key findings of these studies are: 1) the association between symptom overreporting and antisocial features is a) present in self-report measures, but not in behavioral measures, and b) stronger in the punitive setting than in the therapeutic setting; and 2) participants who overreport symptoms a) are prone to attribute blame for their offense to mental disorders, and b) tend to report heightened levels of antisocial traits, but the reverse is not true. In the aggregate, the data do not endorse the use of antisocial features as indicators of assessment validity, on the contrary; they suggest that such practice leads to large proportions of misclassifications.

### Chapter 3

With the study described in Chapter 3, we turned our attention from general antisocial behavior to specific features that may underlie the possible link between antisocial traits and noncredible symptom presentations; namely deceitfulness, lack of remorse, and blame externalization. We found 57 forensic psychiatric inpatients willing to complete an SVT, a self-report instrument measuring lack of remorse and criminal blame externalization, and a semi-structured interview about their most recent offense. To quantify patients' deceitfulness, we checked the information collected via the semi-structured interview against the official records of patients' offenses, and we gathered therapists' and treatment supervisors' judgment on the extent to which patients denied their delinquency.

The results echo those from the two earlier studies: The main finding was that patients who were inaccurate when reporting their delinquency, or who showed little regret for their offenses, or who tended to blame their offenses on external factors, had similar levels of symptom exaggeration as those without these tendencies. As in the first study involving forensic psychiatric inpatients, antisocial personality disorder had no predictive value for noncredible symptom presentations. These results square with previous research into the relationship between antisocial traits and feigned symptomatology. The most important point that can be taken from our studies and the ones that preceded them is that the relationship between antisocial features and noncredible symptom presentations is not sufficiently strong for antisocial features to be useful markers of assessment validity. The presence or absence of antisocial traits should therefore have no bearing on validity assessment.

## Chapter 4

The research detailed in Chapter 4 encompasses a meta-analytic review of all published studies that used the Structured Inventory of Malingered Symptomatology (SIMS) to detect feigned psychopathology. The SIMS is an SVT that presents respondents with 75 pseudo-pathological symptoms that are to be endorsed or rejected via a true-false response format.

The accumulated data show that the diagnostic potential of the SIMS ranges from 59% to 91% correct classifications, and that coaching (i.e., education to pass validity tests while feigning) only has a moderately suppressing effect on SIMS scores. The greater part of the published SIMS data pertain to the traditional cut scores of the SIMS (>14 and >16) and confirm that these are heavily geared toward screening purposes, providing high sensitivity (.91–1.00) yet meager specificity (.37–.70; which corresponds to a false-positive rate of .30–.63). The false-positive risk is exacerbated in patients who suffer from mentally disabling pathology, such as schizophrenia, intellectual disability, and psychogenic non-epileptic seizures.

When reviewing the SIMS qualitatively, two significant shortcomings become apparent. First, the SIMS lacks any means to distinguish symptom overreporting from irrelevant responding (i.e., indiscriminate affirmation or rejection of items). Second, the SIMS does not contain items that are *intended* to cover plausible symptoms, yet it does contain items that *actually do* refer to plausible symptoms. This may well explain the sensitivity of the SIMS to authentic symptomatology, which is discernable in the aggregated data. These two flaws of the SIMS could be amended by incorporating items that tap disengaged responding, and by altering the scoring format to account for items that refer to authentic symptomatology.

Our meta-analysis also yielded two interesting subsidiary findings: Experimental feigners produce SIMS scores that are similar to the scores of real-world feigners; and healthy individuals produce scores that are significantly lower than those of patients. The former finding alleviates concerns about the external validity of experimentally impelled feigning, whereas the latter underscores that healthy individuals cannot be substituted for genuine patients to form control groups.

## Chapter 5

Chapter 5 chronicles our research into a PVT; the Malingering Scale Vocabulary and Abstraction test (MSVA). The MSVA is an atypical PVT in that it does not mimic a memory test, as most PVTs do; rather, it is analogous to vocabulary and abstract reasoning tasks of IQ tests. We translated the MSVA and examined its diagnostic utility by administering it to primary school children ( $n = 41$ ), forensic inpatients with intellectual disability ( $n = 25$ ), forensic inpatients with psychiatric symptoms ( $n = 57$ ), and undergraduate students ( $n = 30$ ,  $n = 79$ , and  $n = 90$ , respectively). In most samples, we employed a simulation design; we assigned participants randomly to a condition in which they were asked to respond honestly or to a condition that prompted them to feign cognitive impairment. In the sample of

forensic inpatients, we took a quasi-known-groups approach, with the SIMS as criterion to segregate participants who responded noncredibly. In the undergraduate samples, we employed the SIMS and two well-validated PVTs (the Amsterdam Short-Term Memory Test; ASTM, and the Test of Memory Malingering; TOMM) as benchmarks to compare the MSVA with.

With cutoffs chosen so as to keep false-positive rates below 10%, detection rates of experimentally feigned cognitive impairment were high in children (90%) and inpatients with intellectual disability (100%), but low in adults without intellectual disability (46%). The rates of significantly below-chance performance were low (4%), except in children (47%) and intellectually disabled inpatients (50%). The reliability of the MSVA was excellent, and its diagnostic efficacy was not significantly affected by forewarning of validity testing or by provision of information about validity testing. The MSVA was about as effective as the TOMM in differentiating between honest and feigning respondents, and slightly less effective than the ASTM. Nevertheless, the MSVA still identified several noncredible respondents that the ASTM and TOMM failed to pick up, indicating that the MSVA has added value, and is worth investigating further. Taken together, our studies are an initiatory step towards the validation of the MSVA as a performance validity test that taps noncredible deficits in vocabulary and abstract reasoning rather than memory.

## Chapter 6

Chapter 6 revolves around an SVT of our own making; the Symptom and Disposition Interview (SDI). By developing the SDI, we explored an unresolved and rarely addressed issue in validity assessment; the differentiation between factitious and malingered symptom presentations. Over and above measuring noncredible symptomatology, the SDI aims to quantify internal incentives that are associated with factitious disorder (i.e., the need for sympathy, attention, and care that comes with the “sick role”). This is effectuated through items that inquire into the willingness to assume the patient role. For example, various items gauge the readiness to undergo additional or unpleasant treatment or diagnostic procedures, to participate in scientific studies or patient support groups, or to vent feelings of indignation for being disregarded, misunderstood, not taken seriously, or otherwise treated incorrectly.

We investigated the qualities of the SDI and compared them with those of a traditional SVT (the SIMS) in two simulation studies, each with factitious and malingering conditions ( $n = 24\text{--}30$  per condition) and a clinical control group ( $n = 34$ ,  $n = 40$ ). The first study yielded more favorable results than the second, yet, overall the results were positive: The SDI was as effective in detecting feigned symptom presentations as the SIMS, and superior in differentiating factitious from malingered symptom presentations. Nonetheless, it is evident that additional research, particularly in clinical samples, is indispensable to appraise the utility of the SDI and the detection strategies upon which it is predicated. Our discourse on the development of the SDI and the associated studies can be viewed as

groundwork that may serve as catalyst for the progress of psychometric assessment of factitious disorder and malingering.

## Chapter 7

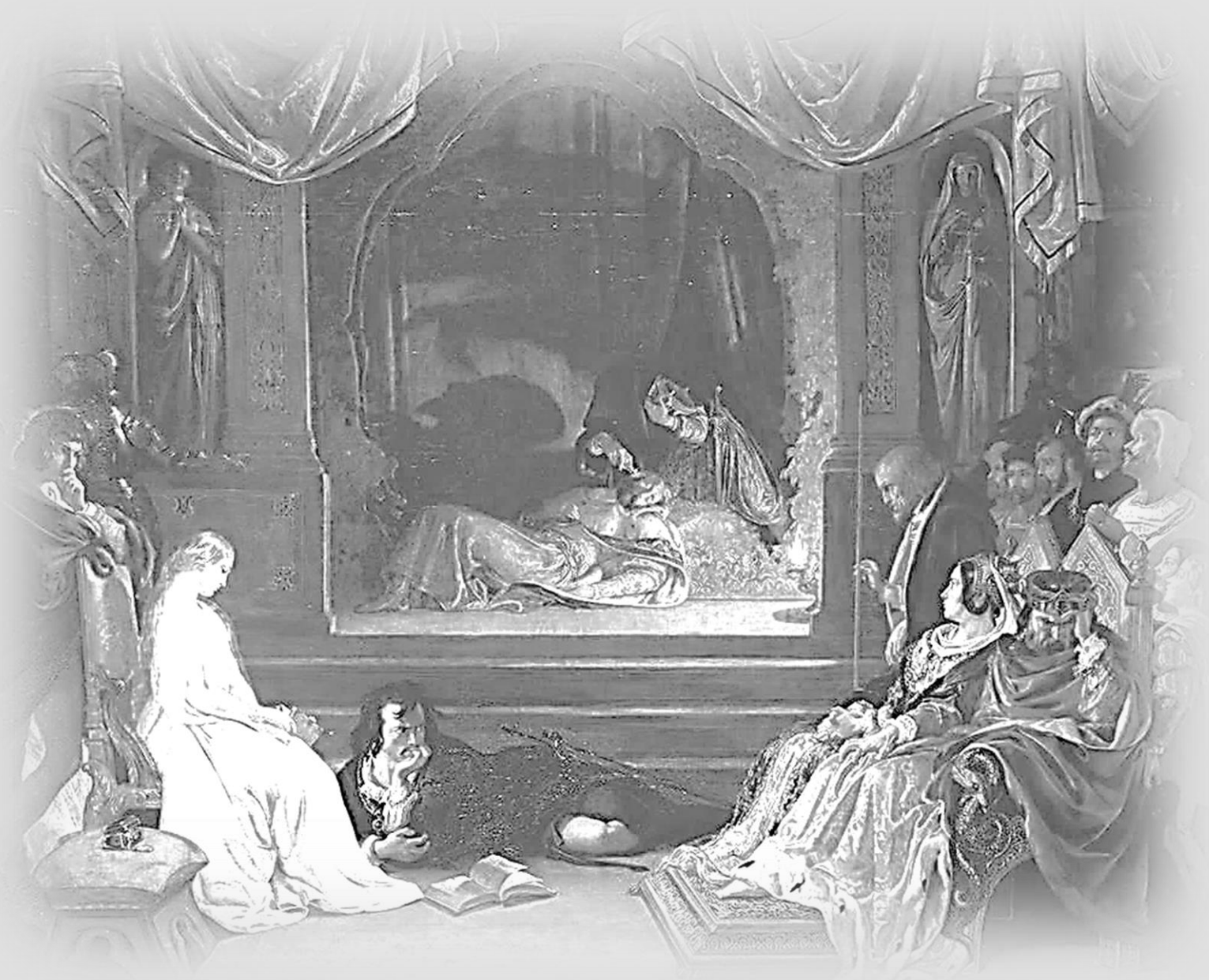
In addition to discussing the main findings and conclusions that are presented in the foregoing chapters, Chapter 7 explicates the likelihood ratio (LR) as an actuarial technique for interpreting and combining validity test data. The LR can be conceived as the degree to which a particular validity test score fits better with the hypothesis that a symptom presentation is noncredible than with the hypothesis that a symptom presentation is credible, or put differently; the degree to which the hypothesis of noncredible symptomatology is more probable after a particular validity test score is obtained compared with how probable it was before the score was obtained.

Every validity test result is associated with a certain sensitivity and specificity value. The LR is calculated by dividing sensitivity by 1 minus specificity ( $LR = \text{sensitivity}/[1 - \text{specificity}]$ ). LRs range from 1 to infinity; the higher the LR above 1, the greater the increase in the pre-test probability of noncredible symptomatology. The pre-test probability is equal to the base rate noncredible symptomatology, or, if other validity tests have already been administered, to the post-test probability of the previous tests. The post-test probability of noncredible symptomatology can be arrived at by first converting the pre-test probability to odds (i.e.,  $\text{pre-test probability}/[1 - \text{pre-test probability}] = \text{pre-test odds}$ ), then multiplying the result with the LR, and then converting the outcome back to probability (i.e.,  $\text{post-test odds}/[1 + \text{post-test odds}] = \text{post-test probability}$ ).

LRs can be “chained” by using the post-test odds from the first test as pre-test odds for the second test, and so forth. When chaining LRs, two caveats must be kept in mind. These are that post-test probabilities of noncredible symptom presentations are overestimated if validity test results are dependent (i.e., correlated) in clinical populations, and if test results are chained without at least two results being associated with high specificity (e.g., specificity  $\geq .90$ ).

In sum, the chaining of LRs is a quantitative method of decision-making that, when employed with validated validity tests, caps off the actuarial approach to validity assessment. The method involves simple calculations and its conclusions follow logically from empirically established knowledge. Thus, it encourages practitioners to work transparently and scientifically, and, in the process, increase diagnostic probability and decrease diagnostic error.

**Samenvatting**  
**(Summary in Dutch)**





Het onderzoek dat in deze dissertatie gepresenteerd wordt, is uitgevoerd om artsen en psychologen beter in staat te stellen om een inschatting te maken van de accuraatheid waarmee patiënten symptomen presenteren. De mate waarin mensen hun symptomatologie accuraat presenteren aan klinici wordt ook wel “symptoomvaliditeit” genoemd. Hoofdstukken 2 en 3 gaan over antisociale gedragskenmerken als voorspellers van gebrekkige symptoomvaliditeit. Hoofdstukken 4, 5 en 6 handelen over psychometrische tests om symptoomvaliditeit te meten (zogenaamde “validiteitstests”). Bij wijze van inleiding wordt in hoofdstuk 1 kort stilgestaan bij de geschiedenis van validiteitstests en bij enkele veel voorkomende mythes en misverstanden over symptoomvaliditeit. Hoofdstuk 7 omvat een discussie van de belangrijkste bevindingen van de studies uit de voorgaande hoofdstukken en beschrijft methodologische overwegingen, klinische implicaties en aanbevelingen voor toekomstig onderzoek. Om de praktische toepasbaarheid van dit proefschrift te vergroten, wordt in hoofdstuk 7 ook uitleg gegeven over de *likelihood ratio* (“aannemelijkheidsquotient”) als kwantitatieve methode om validiteitstest-scores te combineren en te interpreteren.

## Hoofdstuk 1

Hoofdstuk 1 introduceert de drie typen validiteit die tezamen de accuratesse van diagnoses bepalen; testvaliditeit, symptoomvaliditeit en oordeelsvaliditeit. Testvaliditeit is de mate waarin een test datgene meet wat men ermee beoogt te meten. Symptoomvaliditeit is de mate waarin patiënten hun symptomatologie accuraat (d.w.z., open en consciëntieus) presenteren tijdens klinisch onderzoek. Oordeelsvaliditeit is de mate waarin diagnostische conclusies gestoeld zijn op logica en empirische gegevens. Hoofdstuk 1 beschrijft tevens dat besluitvorming in de diagnostiek via twee diametrale methodes plaats kan vinden; de klinische methode en de actuariële methode. Bij de klinische methode worden besluiten genomen of conclusies getrokken op basis van menselijke cognitie (bijv., “klinische intuïtie”); bij de actuariële methode wordt uitsluitend gebruik gemaakt van empirische beslisregels.

De voornaamste mythes en misverstanden die in hoofdstuk 1 de revue passeren zijn het onterechte geloof in de effectiviteit van de klinische methode om symptoomvaliditeit te evalueren; het onjuiste idee dat geveinsde symptomatologie een zeldzaam fenomeen is; de foute aanname dat geveinsde symptomatologie een indicatie is van authentieke symptomatologie; en het verkeerde idee dat antisociale gedragskenmerken samenhangen met het veinzen van symptomen.

Verder beschrijft hoofdstuk 1 dat validiteitstests voortkwamen uit pogingen om functionele neurologische symptomen te detecteren met *two-alternative forced-choice* tests. Tegenwoordig worden twee typen validiteitstests onderscheiden; “prestatievaliditeitstests” (PVTs) en “symptoomvaliditeitstest” (SVTs). PVTs zijn quasi-cognitieve tests en worden gebruikt om de inzet van respondenten tijdens cognitieve taken te bepalen (d.w.z., PVTs meten cognitief onderpresteren). SVTs zijn vragenlijsten of gestructureerde interviews waarin pseudo-symptomatologie is opgenomen om daarmee de accuraatheid van zelf-

rapportage over symptomatologie in te schatten (d.w.z., SVTs meten het overrapporteren van symptomen). De prominentste beperking van validiteitstests is dat ze niet duidelijk kunnen maken *waarom* er bij een psychologische evaluatie mogelijk iets schort aan symptoomvaliditeit; ze kunnen slechts aangeven *of* er iets schort aan symptoomvaliditeit.

## Hoofdstuk 2

Hoofdstuk 2 handelt over twee studies die opgezet zijn om een idee te toetsen dat duidelijk in de *Diagnostic and Statistical Manual of Mental Disorders* ("diagnostisch en statistisch handboek voor psychische stoornissen" 5de ed.) naar voren komt; namelijk dat antisociale gedragskenmerken indicatief zijn voor gebrekkige symptoomvaliditeit (d.w.z., dat antisociale gedragskenmerken samenhangen met het veinzen van symptomen). In de eerste studie namen we bij 57 forensisch-psychiatrische patiënten een SVT en een instrument voor symptoom-minimalisatie (het onderrapporteren van symptomen) af. Daarnaast kwantificeerden we antisociaal en disruptief gedrag van deelnemers door middel van een gedragsobservatie-instrument dat we scoorden op basis van logboeken die door behandelend en verplegend personeel dagelijks werden bijgehouden. In de tweede studie vulden 45 gedetineerde jongeren een SVT en zelf-rapportage instrumenten voor psychopathie en schuld-attributie in. We bepaalden antisociale gedragskenmerken aan de hand van de zwaarte van opgelegde straffen, het al dan niet aangemerkt zijn als "veelpleger", en het aantal opgelegde disciplinaire maatregelen gedurende detentie.

De belangrijkste bevindingen van deze studies zijn: 1) de samenhang tussen het overrapporteren van symptomen en antisociale gedragskenmerken is a) aanwezig in zelf-rapportage instrumenten voor antisociale gedragskenmerken, maar niet in gedragsobservatie-instrumenten, en b) sterker in de punitieve setting dan in de therapeutische setting; en 2) deelnemers die symptomen overrapporteren zijn geneigd om a) schuld voor hun delictgedrag te attribueren aan psychische problemen, en b) meer antisociale gedragskenmerken te rapporteren, maar het tegenovergestelde geldt niet (deelnemers die bovenmatig antisociale gedragskenmerken rapporteren zijn *niet* bovenmatig geneigd om symptomen over te rapporteren). Bij elkaar genomen leveren de data geen bewijs voor het bestaan van een duidelijk verband tussen antisociale gedragskenmerken en symptoomvaliditeit; integendeel: het aannemen van zulk een verband leidt tot grote aantallen misclassificaties van symptoomvaliditeit.

## Hoofdstuk 3

Met de studie die in hoofdstuk 3 beschreven wordt verschuift onze aandacht van algemeen antisociaal gedrag naar de vormen van antisociaal gedrag die het grootste raakvlak hebben met het veinzen van symptomen; namelijk leugenachtigheid, gebrek aan berouw of schuldgevoel, en schuld-externalisatie. We vonden wederom 57 forensisch-psychiatrische patiënten bereid om een SVT en een zelf-rapportage instrument over schuldgevoel en schuld-attributie in te vullen en om vragen te beantwoorden tijdens een semigestructureerd

interview over het indexdelict. Om de leugenachtigheid van deelnemers in te schatten legden we hun antwoorden tijdens het interview over hun indexdelict naast de informatie uit de justitiële documentatie van hun indexdelict en vroegen we naar het oordeel van behandelaren over de mate waarin deelnemers hun indexdelict ontkenden of minimaliseerden.

De resultaten weerspiegelen de uitkomsten van de twee voorgaande studies: de voornaamste bevindingen zijn dat het overrapporteren van symptomen evenveel voorkomt onder patiënten die de neiging hebben om niet accuraat te zijn bij het beschrijven van hun indexdelict, of weinig berouw te tonen over hun indexdelict, of de schuld voor hun delictgedrag te externaliseren, als onder patiënten die zulke neigingen niet vertonen. Zowel in deze als in de eerste studie waaraan forensisch-psychiatrische patiënten deelnamen had de diagnose “antisociale persoonlijkheidsstoornis” geen voorspellende waarde voor het overrapporteren van symptomen. Deze resultaten stroken met de uitkomsten van eerder onderzoek naar de mogelijke samenhang tussen antisociale gedragskenmerken en symptoomvaliditeit: het belangrijkste punt dat voortvloeit uit de bevindingen van de studies op dit gebied is dat de samenhang tussen antisociale gedragskenmerken en symptoomvaliditeit dermate zwak is dat de voorspellende waarde van antisociale gedragskenmerken voor symptoomvaliditeit te verwaarlozen is. De aan- of afwezigheid van antisociale gedragskenmerken zou geen invloed moeten hebben op de evaluatie van symptoomvaliditeit.

#### Hoofdstuk 4

Hoofdstuk 4 is een meta-analyse van alle gepubliceerde studies die de *Structured Inventory of Malingered Symptomatology* (SIMS) gebruikten om het overrapporteren van symptomen te meten. De SIMS is een SVT die bestaat uit 75 stellingen die pseudo-pathologie beschrijven; respondenten dienen bij elke stelling aan te geven of die op hen van toepassing is.

De geaggregeerde data laten zien dat het diagnostische potentieel van de SIMS tussen de 59% en 91% correcte classificaties ligt, en dat coaching (d.w.z., aanwijzingen hoe men kan slagen voor validiteitstests terwijl men symptomen veinst) slechts een beperkt drukkend effect heeft op SIMS scores. Het overgrote deel van de gepubliceerde SIMS data heeft betrekking op de traditionele afkapscores van de SIMS (>14 en >16) en bevestigt dat deze toegespitst zijn op screening; ze genieten hoge sensitiviteit (.91–1.00) maar lage specificiteit (.37–.70; wat overeenkomt met een fout-positief kans van .30–.63). Het risico op fout-positieve classificatie is zelfs nog hoger bij patiënten met beperkte of verstoorde mentale capaciteiten, zoals het geval is bij schizofrenie, verstandelijke beperking, en psychogene niet-epileptische aanvallen.

Een kwalitatieve beschouwing van de SIMS brengt twee majeure tekortkomingen aan het licht. Ten eerste ontbreekt het de SIMS aan mogelijkheden om onderscheid te maken tussen het overrapporteren van symptomen en het willekeurig reageren op items. Ten tweede bevat de SIMS geen items die *bedoeld zijn* om authentieke symptomatologie te meten,

maar wel items die *in zekere mate* aan authentieke symptomatologie refereren. Dit verklaart mogelijk waarom de SIMS, zoals blijkt uit de geaggregeerde data, gevoelig is voor authentieke symptomatologie. Deze twee gebreken van de SIMS zouden geredresseerd kunnen worden door items aan de SIMS toe te voegen die willekeurige antwoordtendensen detecteren en door de score formule aan te passen zodat de items die feitelijk aan authentieke symptomen refereren niet meer meegenomen worden in de bepaling van symptoomvaliditeit.

Onze meta-analyse leverde nog twee interessante bevindingen op: studiedeelnemers die gevraagd worden om symptomen te simuleren produceren SIMS scores die vergelijkbaar zijn met die van échte simulanten, en gezonde deelnemers die gevraagd worden om symptomen accuraat te rapporteren produceren SIMS scores die lager zijn dan die van patiënten. De eerste bevinding vermindert zorgen over de externe validiteit van experimentele simulanten, en de tweede bevinding onderstreept dat gezonde mensen niet thuis horen in klinische controle groepen van experimenteel simulatie-onderzoek.

## Hoofdstuk 5

Het onderwerp van hoofdstuk 5 is een PVT; de *Malingering Scale Vocabulary and Abstraction test* (MSVA). De MSVA is een atypische PVT omdat het *niet*, zoals de meeste PVTs, lijkt op een geheugentest, maar aandoet als vocabulaire- en abstract-redeneren-taken van IQ tests. We vertaalden de MSVA en onderzochten de diagnostische slagkracht ervan in verschillende experimentele studies met kinderen ( $n = 41$ ), forensische patiënten met een verstandelijke beperking ( $n = 25$ ), forensisch-psychiatrische patiënten ( $n = 57$ ), en studenten (respectievelijk  $n = 30$ ,  $n = 79$  en  $n = 90$ ). In het merendeel van de groepen hanteerden we een experimentele opzet; we verdeelden deelnemers willekeurig over een conditie waarin gevraagd werd om optimaal te presteren en een conditie waarin getracht werd om deelnemers cognitieve beperkingen te laten veinzen. Bij de forensisch-psychiatrische patiënten gebruikten we de SIMS om de groep te splitsen in deelnemers die mogelijk cognitieve beperkingen veinsden en deelnemers die dat niet deden. In de groepen studenten zetten we naast de SIMS tevens twee gevalideerde maten voor cognitief onderpresteren in (de Amsterdamse Korte Termijn Geheugen test; AKTG, en de *Test of Memory Malingering*; TOMM) om de MSVA mee te vergelijken.

Met afkappunten waarbij het risico op fout-positieve classificatie minder dan 10% is, was de detectie ratio van experimenteel gesimuleerde cognitieve beperkingen hoog bij kinderen (90%) en forensische patiënten met een verstandelijke beperking (100%), maar laag bij volwassenen zonder verstandelijke beperking (46%). Het aantal gevallen van prestatie onder kansniveau was klein (4%), behalve bij kinderen (47%) en forensische patiënten met een verstandelijke beperking (50%). De betrouwbaarheid van de MSVA was goed en de diagnostische effectiviteit werd niet significant ondermijnd door waarschuwingen voor validiteitstest of door het verstrekken van informatie over validiteitstests. De MSVA was ongeveer even effectief als de TOMM in het maken van onderscheid tussen eerlijke en

simulerende deelnemers, en iets minder effectief dan de AKTG. Desalniettemin detecteerde de MSVA verscheidene simulerende deelnemers die door de AKTG en de TOMM over het hoofd werden gezien; dit geeft aan dat de MSVA toegevoegde waarde heeft en daarom verder onderzocht zou moeten worden. Bij elkaar genomen vormen onze studies een eerste stap richting de validatie van de MSVA als PVT die zich richt op gesimuleerde beperkingen in vocabulaire en abstract redeneren in plaats van geheugen.

## Hoofdstuk 6

Hoofdstuk 6 draait om een SVT die we zelf ontwikkeld hebben; de *Symptom and Disposition Interview* (SDI). Met de SDI exploreerden we een onopgelost en nauwelijks onderzocht probleem in de evaluatie van symptoomvaliditeit; namelijk het maken van onderscheid tussen simulatie en nagebootste stoornissen ("malingering" en "factitious disorder"). Bij simulatie veinzen mensen symptomen omwille van externe voordelen; ze trachten er tastbaar gewin mee te behalen, zoals financiële toekenningen of het ontlopen van plichten of verantwoordelijkheden; bij nagebootste stoornissen wordt het veinzen van symptomen gedreven door interne motieven; bijvoorbeeld om aandacht, hulp of sympathie te verkrijgen. In tegenstelling tot simulatie worden nagebootste stoornissen beschouwd als psychiatrische aandoeningen. Naast het detecteren van symptoom-overrapportage dient de SDI om verschillende interne motieven te peilen die geassocieerd worden met nagebootste stoornissen. Dit wordt bewerkstelligd door items die aftasten in hoeverre respondenten bereid zijn patiënt-gerelateerde activiteiten te ondernemen; bijvoorbeeld om niet-essentiële of onprettige procedures te ondergaan ter diagnose of behandeling; om deel te nemen aan praatgroepen of wetenschappelijke studies; of om frustratie te uiten over niet begrepen of serieus genomen of anderszins niet correct behandeld worden.

We onderzochten de kwaliteiten van de SDI en vergeleken deze met die van een traditionele SVT (de SIMS) in twee experimentele studies; elk met simulatie- en nagebootste stoornis-condities ( $n = 24-30$  per conditie) en een klinische controle groep ( $n = 34, n = 40$ ). De eerste studie leverde mooiere resultaten op dan de tweede, maar over het geheel genomen waren de resultaten positief. De SDI was even effectief als de SIMS in het detecteren van symptoom-overrapportage en superieur in het differentiëren tussen de simulatie- en nagebootste stoornis-condities. Uiteraard is verder onderzoek, in het bijzonder in klinische populaties, nodig om de kenmerken van de SDI en de effectiviteit van de gebruikte detectie-strategieën beter in kaart te brengen. Ons onderzoek naar de SDI kan het best beschouwd worden als exploratief voorwerk dat als basis kan dienen voor vervolgonderzoek naar de psychometrische evaluatie van simulatie en nagebootste stoornissen.

## Hoofdstuk 7

Naast een discussie van de belangrijkste resultaten en conclusies van de studies die beschreven werden in voorgaande hoofdstukken biedt hoofdstuk 7 een uiteenzetting van de *likelihood ratio* (LR) als kwantitatieve methode om validiteitstestcores te combineren en te

interpreteren. De LR kan gezien worden als de mate waarin een bepaalde validiteitstestscore meer strookt met de hypothese dat een symptoompresentatie *niet* accuraat is dan met de hypothese dat een symptoompresentatie *wel* accuraat is. Anders gezegd kan de LR gezien worden als de mate waarin de hypothese dat een symptoompresentatie niet accuraat is meer aannemelijk wordt *nadat* een bepaalde validiteitstestscore bekend is.

Elke validiteitstestscore is geassocieerd met een bepaalde sensitiviteit en specificiteit; sensitiviteit is de kans op een positieve testuitslag bij positieve respondenten en specificiteit is de kans op een negatieve testuitslag bij negatieve respondenten. De LR wordt berekend door sensitiviteit te delen door 1 minus specificiteit ( $LR = \text{sensitiviteit} / [1 - \text{specificiteit}]$ ). LRs variëren van 1 tot oneindig; hoe verder boven 1, hoe groter de toename in waarschijnlijkheid dat een symptoompresentatie niet accuraat is.

De pre-test kans dat een symptoompresentatie niet accuraat is, is gelijk aan de prevalentie van niet-accurate symptoompresentaties, of, indien er reeds validiteitstests zijn afgenomen, aan de post-test kans behorende bij de afgenomen validiteitstests. De post-test kans dat een symptoompresentatie niet accuraat is, wordt berekend door eerst de pre-test kans om te rekenen naar *odds* ("kansverhouding");  $\text{pre-test kans} / [1 - \text{pre-test kans}] = \text{pre-test odds}$ . De pre-test odds kan vervolgens vermenigvuldigd worden met de LR, waarna de uitkomst weer terug kan worden gerekend naar kans via  $\text{post-test odds} / [1 + \text{post-test odds}]$ .

Het combineren van testresultaten wordt bewerkstelligd door het aaneenschakelen ("*chaining*") van LRs: dit houdt in dat de post-test odds van de eerste test gebruikt worden als pre-test odds van de tweede test, enzovoorts. Bij het aaneenschakelen van LRs dienen twee zaken in het oog gehouden te worden; dat de post-test kans op niet-accurate symptoompresentaties overschat wordt indien de resultaten van validiteitstests samenhangen (d.w.z., met elkaar correleren) in klinische populaties, en indien testresultaten aaneengeschakeld worden zonder dat er tenminste twee resultaten geassocieerd zijn met een hoge specificiteit (d.w.z.,  $\geq .90$ ).

Samenvattend kan gesteld worden dat het aaneenschakelen van LRs een kwantitatieve methode is om diagnostische beslissingen te nemen en dat deze methode de actuariële benadering van symptoomvaliditeitstoetsing completeert. De berekeningen zijn eenvoudig (zie blz. 160, tabel 7.2) en de beslissingen die ermee genomen worden stoelen volledig op logica en empirische kennis; kortom, het gebruik van LRs bevordert de transparantie en de wetenschappelijke insteek van diagnostiek en helpt om de diagnostische foutenmarge te reduceren.

