

Personal Problems Questionnaire (PPQ): Normative Data and Utility in Assessing Acquired Neurological Impairment

Martin D. van den Broek¹, Linda Monaci², Jared G. Smith^{3,*}

¹*The Neuropsychology Clinic, Parallel House, Guildford, UK*

²*Private Practice, London, UK*

³*Population Health Research Institute, St. George's, University of London, London, UK*

*Corresponding author at: Population Health Research Institute, St George's, University of London, Cranmer Terrace, Tooting, London SW17 0RE, UK.
E-mail address: jasmith@sgul.ac.uk (J.G. Smith).

Editorial Decision 29 July 2018; Accepted 3 August 2018

Abstract

Objective: The Personal Problems Questionnaire (PPQ) is a measure designed to assess acquired cognitive, emotional, and physical complaints. The present study sought to develop a normative database to allow clinicians and researchers to assess self-reported complaints among people with disabilities, and evaluate the response consistency and validity of their self-report.

Method: 404 community-dwelling participants (*n* 200 males, 204 females) completed the PPQ, as well as an acquired brain injury (ABI) group (*n* 59), mainly following stroke and traumatic brain injuries, and seen for clinical (i.e., non-forensic) evaluations. Multiple regression analyses were conducted to derive norms from the healthy community sample taking into account age, gender, and educational level.

Results: Normative *T* scores and cut-off points for the Clinical and Validity scales were derived, respectively, and used to assess the responses of the ABI group. The results indicated that the ABI group showed good response consistency and elevated scores on the Clinical scales, indicating that the PPQ is likely to be useful in detecting acquired disabilities. On the other hand, scores on the Validity scales were not elevated, indicating that the measures were unaffected by the ABI participants' cognitive difficulties.

Conclusions: The PPQ provides a comprehensive assessment of complaints and response validity and the present study provides further data to assist with its use and interpretation.

Keywords: Assessment; Malingering; Symptom validity testing; Traumatic brain injury; Normative studies; Everyday functioning

Introduction

The Personal Problems Questionnaire (PPQ: [van den Broek, Monaci, & Smith, 2012](#)) is an instrument designed to assess patients' self-identified cognitive, emotional and physical complaints. It comprises 156 questions and 12 Clinical subscales relating to common difficulties in disabled populations. Four Clinical subscales relate to emotional complaints including Anger, Anxiety, Depression, and Stress, four to cognitive limitations relating to Memory, Concentration, Language, and Communication, and four to physical issues, that is, Pain, Neurological Complaints, Somatic complaints, and Disability and Restrictions. In addition, there are three Total Clinical scales that are composite scales for each of the clinical domains (see [Table 1](#)). Patients are presented with a series of questions relating to each type of complaint and asked to rate the degree to which they have experienced the problem in the preceding month. The PPQ was not developed to represent a particular factor structure at the level of the scale themselves, but essentially follows clinical reasoning about physical, emotional, and cognitive complaints, and atheoretical diagnostic criteria, in particular, the DSM system ([American Psychiatric Association, 2013](#)).

A key issue when assessing self-report is determining whether an individual's responses are valid with an international consensus having emerged that assessing response validity should be a feature of clinical assessments ([American Academy of Clinical Neuropsychology Board of Directors, 2007](#); [Bush et al., 2005](#); [Heilbronner et al., 2009](#); [McMillan et al., 2009](#);

Table 1. Clinical subscales and symptoms and behavioral tendencies associated with each scale (adapted from van den Broek et al., 2012). Each subscale comprises 10 items, yielding 40 items for each Total Clinical scale. Each Validity scale comprises 12 unique, non-overlapping items yielding a total of 36 items. The Consistency (Con) scale comprises nine item-pairs

| | |
|-------------------------------------|--|
| <u>Emotional subscales</u> | |
| 1. Anger | Irritability, frustration, verbal & physical aggression, impatience |
| 2. Anxiety | Anxiety, feelings of worry, low self-confidence, fearfulness |
| 3. Depression | Low mood, tearfulness, loss of enthusiasm, anhedonia |
| 4. Stress | Inability to relax, feeling overwhelmed, racing thoughts |
| Total Clinical Emotional Scale | |
| <u>Cognitive subscales</u> | |
| 1. Memory | Everyday forgetfulness, prospective lapses, poor retrograde recall |
| 2. Concentration | Distractibility, difficulty maintaining attention, focusing, multitasking |
| 3. Language & Communication | Word finding, garrulousness, tangentially, difficulty with understanding |
| 4. Executive | Carelessness and disorganization, impulsiveness, inappropriate responses |
| Total Clinical Cognitive Scale | |
| <u>Physical subscales</u> | |
| 1. Pain | Pain impeding life, use of analgesia, perceived pain-related restrictions |
| 2. Neurological | Motor restrictions, diminished sensation, dizziness, visual problems |
| 3. Somatic | Sensitivity to illness, use of specialists, gastrointestinal, bowel complaints |
| 4. Disability & Restrictions | Reliance on carers, need for aids, self-care and mobility restrictions |
| Total Clinical Physical Scale | |
| <u>Validity scales</u> | |
| 1. Response Consistency (Con) | |
| 2. Validity Emotional Scale (vEmot) | |
| 3. Validity Cognitive Scale (vCog) | |
| 4. Validity Physical Scale (vPhys) | |

van den Broek and Sembi, 2017). Non-credible presentations are a feature of those with psychogenic non-epileptic attacks, pain, somatization, and medically unexplained symptoms (Cragar, Berry, Fakhoury, Cibula, & Schmitt, 2006; Etherton, Bianchini, Ciota, Heinly, & Greve, 2006), as well as those involved in litigation (Greve et al., 2006; Larrabee, 2003; Young, 2014). Patients may present with either under-performance on cognitive tests or over-reporting on self-report measures or a combination of the two. A wide range of stand-alone performance validity tests (PVTs) are now available to delineate under-performance, such as the Word Memory Test (Green, 2003) and the Test of Memory Malingering (Tombaugh, 1996), as well as embedded measures like those included in the Advanced Clinical Solutions battery (Holdnack, Whipple Drozdick, Weiss & Iverson, 2013). Over-reporting can be delineated using questionnaires like the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008), the Personality Assessment Inventory (PAI; Morey, 1991), and the Neurobehavioral Symptom Inventory (Bodapati et al., 2018; Meyers, English, Miller & Lee, 2015), with elevated scores on validity scales potentially indicating exaggerated complaints.

van den Broek and colleagues (2012) pointed out that PVTs focus on cognitive issues alone, particularly memory problems, and so do not address the full range of complaints (cognitive, emotional, and physical) that are potentially reported by disabled individuals, particularly those with acquired brain injuries. On the other hand, questionnaires tend to have psychopathology as their primary focus and rely on ascertaining the degree to which complaints are endorsed. Exaggeration or over-reporting is inferred when patients endorse an excessive number of complaints both exceeding that of healthy normal subjects and appropriate patient comparison groups. However, interpretation can be complicated as elevated scores can also be found in those with severe, but genuine psychopathology, so attenuating the utility of the method (Ben-Porath & Tellegen, 2008). Multidimensional questionnaires can also be taxing to complete and involve lengthy administration times which affect their utility in some settings, such as busy outpatient clinics where briefer measures may be preferable and more acceptable to patients. One such measure that screens for distorted symptoms is the Self-Report Symptom Inventory (SRSI; Merten, Merckelbach, Giger, & Stevens, 2016). However, in-depth analysis of the genuine symptom scales is lacking to date, rendering findings concerning symptom validity difficult to interpret.

van den Broek and colleagues (2012) previously evaluated the utility of the PPQ in addressing some of these limitations. The PPQ incorporates concepts developed by Rogers (2008) who outlined two broad strategies that can be used to identify people who present with non-valid symptoms: amplified presentations and unlikely presentations. The first involves monitoring possible over-reporting of symptom frequency, duration and severity of symptoms through the indiscriminate endorsement of problems, and the second involves monitoring for symptoms that are rarely reported by honest responders, symptoms that are implausible, and improbable symptom combinations, that is, symptoms that are common, but rarely occur together. These strategies generally monitor for negative symptom distortion as both involve the endorsement of symptoms (Rogers, 2008).

For each PPQ Clinical domain, there is a corresponding Validity scale of unique, non-overlapping items to check the validity of the patients' self-report (see Table 1). They comprise items that are implausible, overly specified, overly severe, or involve unusual symptom combinations that would not be endorsed by normal individuals, those with a neurological disability or any other disorder. van den Broek and colleagues (2012) examined the utility of the PPQ in differentiating between patients with acquired brain injuries, healthy community-dwelling subjects, and control subjects asked to simulate an acquired brain injury. The Clinical scales satisfactorily differentiated between the groups and the Validity measures had good sensitivity (Cognitive Validity 0.70; Emotional Validity 0.90; Physical Validity 0.90) and high specificity (Cognitive Validity 0.97; Emotional Validity 0.95; Physical Validity 0.98) when detecting simulated complaints.

The present investigation sought to further advance the utility of the PPQ in a number of respects. First, the aim was to develop a normative database for the PPQ by recruiting a representative sample of healthy normal adults living in the community, so allowing clinicians to compare patients' responses on the Clinical and Validity scales against those found in the normal population. Second, a measure of response consistency was developed to check that semantically similar items in the questionnaire were responded to in a consistent manner. Third, the role of demographic factors that potentially influence responding on the PPQ was also examined. Fourth, a group of patients with acquired brain injuries (ABI) were studied to provide additional confirmatory support for the clinical use of the measure. In particular, it was anticipated that the ABI group would differ from community subjects on the Clinical Scales but, for the method to be useful, they would not differ in their endorsement of items on the Validity and Consistency scales when appropriate cut-off scores were set. Finally, with the development of a normative database, it was intended to provide guidelines for the interpretation of the questionnaire in clinical practice.

Method

Participants

Healthy community-dwelling participants (Community group) were included in the study provided they were able to understand and read English and complete the PPQ. Exclusion criteria were as follows: uncorrected visual impairment; undergoing treatment for alcohol/drug dependency; consuming ≥ 3 alcoholic drinks on two or more nights per week; having sought attention from a professional for memory or cognitive problems; having a history of traumatic brain injury involving loss of consciousness for ≥ 5 minutes and/or requiring hospitalization > 24 h; any other medical or psychiatric condition that could affect cognitive functioning (e.g., stroke, epilepsy, brain surgery, encephalitis, meningitis, multiple sclerosis, Parkinson's disease, Huntington's disease, Alzheimer's disease, schizophrenia, or bipolar disorder); currently receiving treatment for a psychiatric condition.

A total of 404 community-dwelling participants (200 males, 204 females) were recruited via opportunistic sampling over a 3-year period (2009–12). This was more than the suggested (minimum) criterion of 400 total participants needed for a sufficiently precise estimate of population reliability (Charter, 1999), although the regression-based norming procedure adopted here typically requires a smaller sample than traditional discrete-based methods to obtain equally precise norms (Oosterhuis, van der Ark, & Sijtsma, 2016). Participants were sampled so that the age range and gender were representative of the UK population and all participated on a voluntary basis. Their mean age was 40.7 years ($SD = 14.3$, range = 18–69). Most participants (250; 62%) were educated to a high level (i.e., Diploma or Degree qualification), 122 (30.2%) had completed school-level qualifications only (General Certificate of Secondary Education and/or A-Level; US equivalent Grade 10–12) and 31 (8%) had no school qualifications. In addition, 59 participants with acquired brain injuries (ABI; 37 males, 22 females) were recruited from inpatient and outpatient clinics of a Regional Neuropsychology service. The distribution of ages in the ABI group was comparable to Community participants (mean age was 43.6 years, $SD = 14.8$, range = 20–69; $p = .160$), although there was a trend for ABI participants to be less well-educated (high-level education $n = 28$ or 47.5%; school-level qualifications only $n = 26$ or 44%; no school qualifications $n = 5$ or 9%; $p = .086$).

The ABI group comprised patients with traumatic brain injuries ($n = 22$ or 37%; 12 severe, five moderate, five mild; defined by DSM five criteria; American Psychiatric Association, 2013), cerebral vascular accident ($n = 28$ or 49%), encephalitis ($n = 3$), multiple sclerosis ($n = 3$), cerebral tumor ($n = 2$), and hydrocephalus ($n = 1$). Exclusion criteria were as follows: uncorrected visual impairment; significant cognitive impairment precluding administration of the PPQ; undergoing treatment for alcohol/drug dependency; current treatment for severe mental illness; involvement in personal injury litigation.

The study was approved by the Ethical Committee of Wandsworth National Research Ethic Service (NRES; Reference number 08/H0803/138) and the Research and Development Office, St. George's Hospital, London.

Measures

After obtaining informed consent to participate, demographic details were collected on each subject (age, sex, educational qualifications) and, in the case of the ABI group, the patient's diagnosis was extracted from the clinical record. Participants were asked to read the PPQ instructions which indicated that they should respond to the questions by indicating the degree to which they had experienced each problem during the preceding month on a three point rating scale: *Never* (scored 0), *Sometimes* (scored 1), and *Often* (scored 2). The scores on the 12 Clinical Scales, therefore, ranged between 0 and 20 and on the Validity Scales from 0 to 24. An index of variable response consistency (comparable to VRIN in the MMPI family of tests: Ben-Porath & Tellegen, 2008; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; Hathaway & McKinley, 1943) was also developed by matching nine pairs of items that were semantically similar in content (i.e., the construct the items were written to measure). Discrepant responses on each of the pairs were scored by taking the absolute difference between item scores.

Statistical analysis

Internal consistency of the PPQ scales and subscales was assessed separately for control and ABI participants using Cronbach's alphas calculated from participants who provided responses to all items on the relevant (sub)scale. With respect to response consistency, differences for all index pairs in the control group were summed and a *T* score derived from the (raw) inconsistency total scores. Subsequently, the *T* scores were applied to the raw total scores of the ABI group. A *T* score of 80 or higher was adopted to indicate excessive variable response inconsistency, in line with comparable measures (e.g., MMPI-2-RF VRIN; Ben-Porath & Tellegen, 2008).

To derive norms from the healthy sample, multiple regression analyses were conducted. Scores on the PPQ scales were the dependent variables whereas age, gender, and educational level were included as independent variables. Because scores on all three Clinical scales (Cognitive, Emotional, Physical) had positively skewed distributions, scores were transformed by taking the square root before entering into linear regression analyses. The distribution of the Validity scale (Cognitive, Emotional, Physical) scores were extremely skewed, restricted to a small part of the possible range of values (80% of Community participants scored 0, 1, or 2 on each scale). Consequently, the risk of violating the normality of residuals assumption was high using linear regression (Fox, 1997). As such, scores were dichotomized according to a "likely non-credible" classification (based on a cut-off score that yielded approximately 90% specificity) and analyzed using binary logistic regression. In all models, dummy coding was used for categorical predictors: gender was coded with female = 1 and male = 0; educational level (Low = none; Medium = GSEs, A Levels, or Other; US equivalent Grade 10–12; High = Diploma or Degree) was coded with two dummies (Low and High) and Medium as the reference category. Linear and quadratic terms were included for the quantitative predictors of age to examine linear and curvilinear age-trajectories (in years minus the mean of 40.72, to prevent collinearity between the linear and quadratic term). In the absence of hypotheses about interaction between demographics variables, no interactions were included. Each model was reduced in a stepwise way by eliminating the least significant predictor if its two-tailed *p* value was above .05. The dummy variables Education—Low and Education—High were always both included if either was significant, although a predictor was not removed from the model if it was included in a higher order term (e.g., Age and Age²). A number of model checks were carried out (Parmenter, Testa, Schretlen, Weinstock-Guttman, & Benedict, 2010; Roelofs et al., 2011; van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006). For regression models of Clinical scales, homoscedasticity was evaluated by grouping participants into quartiles of the predicted scores and applying the Levene test to the residuals. Kolmogorov–Smirnov tests were administered on the residual values to establish that they approximately followed a normal distribution. The occurrence of multicollinearity was checked by calculating the Variance Inflation Factors (VIFs) in regression models of Clinical scales (Belsley, Kuh, & Welsch, 1980) and standard errors for beta coefficients in regression models of Validity scales. Calculation of Cook's distances was performed to identify possible influential cases in both model types (Cook & Weisberg, 1982). Consequently, a regression model was developed for each Clinical index through which it was possible to compute demographically predicted scores. The raw residual or prediction error was calculated, which was the difference between each participant's observed and predicted scale score. Next, participants' standardized residuals were computed by dividing raw residuals by the standard deviation of the control group's raw residuals. Finally, the standardized residuals (*z* scores) were transformed into *T* scores so that each test had a mean of 50 and an *SD* of 10. For the Validity scales, the final logistic models were used to determine independent variables that affected scores and produce normative cumulative frequency tables for each relevant subgroup.

The multiple regression equations derived from healthy controls were subsequently applied to compute demographically predicted Clinical scale scores for the ABI participants. These were subtracted from observed scores and standardized then

converted to *T* scores in the manner described earlier. Independent-group *t*-tests were administered to compare scores with the Community group (homogeneity of variance was examined using Levene's tests) and chi-square tests used to compare proportions scoring abnormally high (i.e., *T* score ≥ 65). Cumulative frequency tables for each Validity scale were calculated for ABI participants and the proportion classified as "likely non-credible" compared with the Community group controlling for independent variables that affected scores in the latter group. The threshold for statistical significance in group comparisons and regression analyses was $p < .05$. All statistical analyses were completed with the Statistical Package for the Social Sciences, Release 22.0 (SPSS, IBM).

Results

Summary Data for Clinical and Validity Scales of the PPQ

The mean Total Clinical index (the sum of the four subscales in each of the three domains), Clinical scale and Validity scale scores for the Community sample are reported in Table 2. Generally, scores on subscales were low, reflecting few reported problems within the sample. Clinical scale totals tended to be higher for cognitive and emotional domains compared with physical problems ($p < .001$). Overall, the Clinical scales showed strong internal consistency of comprising items ($\alpha > .90$). The subtests also evidenced adequate internal reliability (range .67–.83, median = .78). Consistent with the previous simulation study using the PPQ (van den Broek et al., 2012), scores on Validity scales were very low, confined predominantly to 0, 1, or 2 on each scale. The Validity scales also showed adequate internal consistency ($\alpha > .70$). Notably, response consistency was high; the mean score on the index (0–18) was 2.44 ($SD = 1.94$) with 75% of scores between 0 and 4. Only one control participant (i.e., 0.25% of the sample) had an inconsistency *T* score of more than 80.

Previous research on symptom validity and malingering has established a convention that the specificity of the measure should be approximately $\geq 90\%$ (Boone, 2009; Larrabee, 2007; Morgan & Sweet, 2009). The cumulative frequency table

Table 2. Summary statistics for community participants on the PPQ Clinical scales ($n = 404$)

| Clinical domain | Subscale (0–20) | Mean (<i>SD</i>) | Range | α (95% CI) |
|------------------------------|---------------------------|--------------------|-------|-------------------|
| Cognitive | Language & Communication | 3.49 (2.96) | 0–15 | 0.76 (0.73,0.79) |
| | Memory | 3.82 (3.24) | 0–17 | 0.81 (0.78,0.83) |
| | Concentration | 3.85 (3.44) | 0–15 | 0.83 (0.80,0.85) |
| | Executive | 3.58 (3.08) | 0–15 | 0.77 (0.74,0.81) |
| Emotion | Anger | 4.41 (3.40) | 0–16 | 0.81 (0.78,0.84) |
| | Anxiety | 3.23 (2.99) | 0–15 | 0.79 (0.75,0.82) |
| | Depression | 2.74 (3.00) | 0–19 | 0.82 (0.80,0.85) |
| | Stress | 3.96 (3.18) | 0–15 | 0.79 (0.76,0.82) |
| Physical | Pain | 2.84 (2.82) | 0–17 | 0.77 (0.74,0.80) |
| | Neurological | 1.68 (2.05) | 0–11 | 0.67 (0.62,0.71) |
| | Disability & Restrictions | 1.34 (2.00) | 0–13 | 0.71 (0.67,0.75) |
| | Somatic | 2.14 (2.37) | 0–15 | 0.72 (0.68,0.76) |
| Clinical scale totals (0–80) | | | | |
| | Cognitive | 14.75 (11.32) | 0–55 | 0.93 (0.92,0.94) |
| | Emotional | 14.34 (10.79) | 0–49 | 0.93 (0.92,0.94) |
| | Physical | 8.00 (8.01) | 0–56 | 0.91 (0.89,0.92) |
| Validity scale totals (0–24) | | | | |
| | Cognitive | 1.25 (2.16) | 0–14 | 0.76 (0.73,0.80) |
| | Emotional | 1.40 (2.08) | 0–14 | 0.71 (0.66,0.75) |
| | Physical | 1.08 (1.89) | 0–11 | 0.74 (0.71,0.78) |

In the few cases (<1%) where a participant failed to respond to one item on a subscale, the total subscale score was calculated by imputing the mean item score for the missing item; α = Cronbach's alphas; CI = confidence intervals (calculated using Feldt's (1965) formula).

Table 3. Cumulative percentage of community participants scoring at or selected raw score cut-offs on the Validity scales of the PPQ ($n = 404$)

| Validity scale totals | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 12 | 14 |
|-----------------------|------|------|------|------|------|------|------|------|------|------|-----|
| Cognitive (0–24) | 56.2 | 74.3 | 82.2 | 86.6 | 91.6 | 94.3 | 96.3 | 98.8 | 99.3 | 99.5 | 100 |
| Emotional (0–24) | 50.5 | 66.8 | 79.5 | 87.6 | 91.1 | 93.8 | 96.0 | 98.8 | 99.5 | 99.8 | 100 |
| Physical (0–24) | 56.9 | 77.7 | 85.9 | 90.1 | 92.6 | 95.3 | 96.8 | 98.8 | 99.5 | 100 | |

Table 4. Final multiple linear regression models for the Clinical scale totals ($n = 403$)

| Clinical scale totals ($\sqrt{}$) | B | SE B | <i>t</i> | β | R^2 | MSE |
|-------------------------------------|--------|--------|-----------|---------|-------|-------|
| Cognitive | | | | | | |
| (constant) | 2.638 | 0.186 | 14.197*** | | | |
| Age | -0.026 | 0.006 | -4.703*** | -0.237 | | |
| Age ² | 0.002 | <0.001 | 4.521*** | 0.226 | | |
| Gender | 0.453 | 0.149 | 3.048** | 0.145 | | |
| Education—Low | 1.010 | 0.306 | 3.299** | 0.172 | | |
| Education—High | 0.315 | 0.167 | 1.888 | 0.098 | 0.096 | 2.219 |
| Emotional | | | | | | |
| (constant) | 2.704 | 0.185 | 14.640*** | | | |
| Age | -0.018 | 0.005 | -3.290** | -0.168 | | |
| Age ² | 0.002 | <0.001 | 4.074*** | 0.206 | | |
| Gender | 0.370 | 0.148 | 2.501** | 0.121 | | |
| Education—Low | 1.035 | 0.304 | 3.400*** | 0.180 | | |
| Education—High | 0.265 | 0.166 | 1.599 | 0.084 | 0.070 | 2.193 |
| Physical | | | | | | |
| (constant) | 1.809 | 0.151 | 11.948*** | | | |
| Age | 0.002 | 0.005 | 0.394 | 0.020 | | |
| Age ² | 0.002 | <0.001 | 4.465*** | 0.222 | | |
| Education—Low | 1.346 | 0.277 | 4.860*** | 0.252 | | |
| Education—High | 0.330 | 0.151 | 2.181* | 0.113 | 0.101 | 1.820 |

Data regarding education was missing for one female participant; hence the model includes 403 participants only; Coding of the predictors: age = calendar age - 40.72; age² = (calendar age - 40.72)²; gender: female = 1, male = 0; Education: Education—Low = 1; average or high education = 0; Education—High: high education = 1; average or low education = 0. SE = standard error; β = standardized beta; R^2 = adjusted R^2 ; MSE = mean squared error (residual). * $p < .05$; ** $p < .01$; *** $p < .001$

(Table 3) indicates that a cut-off score of ≥ 4 yields a satisfactory level of specificity and ≥ 5 provides an excellent level and so the latter was adopted as a cut-off for the classification of likely non-credible.

Predictors of PPQ Scores and Normative Data

The linear regression models for the Total Cognitive, Emotional, and Physical Clinical scales are presented in Table 4. Skewness and kurtosis values ($>-.2$ and $<.2$) indicated an approximately normal distribution of the residuals of the normative sample; the normality assumption was specifically tested with Kolmogorov–Smirnov tests and not rejected ($p > .01$ for all Clinical scales). The homoscedasticity assumption was satisfied; Levene tests applied to the residual variance of sample quartiles yielded nonsignificant differences for each Clinical scale ($p > .458$). Therefore, the overall residual standard deviation was used to compute participants' standardized residuals. The maximum VIF for the final models was 1.209 indicating no multicollinearity and there was no serious influence of outliers (maximum Cook's distance for the final models equaled .057).

The models show that female participants evidenced higher scores than males on Cognitive and Emotional scales. In addition to linear age effects on Cognitive and Emotional scales, suggesting younger participants indicated greater difficulties in these domains, there was a marked quadratic effect of age on all scales, reflecting higher scores for the youngest (18–29 years) and oldest (60–69 years) participants. Participants with low levels of education also reported more difficulties across all domains. The raw scores on each of the Total Clinical scales were converted to *T* scores based on the regression-based norming procedure described earlier. As an example, consider a 38 year old female with a high level of education (Diploma or Degree qualification) who scores 13 on the Total Emotional scale. This yields a *T* score equivalent of 51.36¹ which is very close to the mean accounting for their sociodemographic profile. On the other hand, if a woman of the same age, but with a low level of education, obtained an Emotional scale score of 13, the calculated *T* score would be 46.19, slightly less than half a standard deviation below the mean, revealing the influence of low education level in the observed model.

Although methodologically advantageous, the approach adopted here to derive normative data—based on a multi-step procedure using regression equations to calculate *T* scores—is not always user-friendly for typical users of an instrument (Van Breukelen & Vlaeyen, 2005). The provision of normative data tables can increase both accessibility and utility of derived

¹ Table 4 gives a (square root) prediction score for the Emotional Scale of 3.403 [$2.704 + (-0.018)(38-40.72) + 0.002((38-40.72)^2) + 0.370(1) + 1.035(0) + 0.265(1)$]. The residual, therefore, equals 0.203 ($\sqrt{13 - 3.403}$) and divided by the standard deviation of the residual (1.490 or $\sqrt{2.219}$) yields a *z* score of 0.136, which equals a *T* score of 51.36.

norms (Van der Elst et al., 2006). Supplementary Tables 1–3 provide normative tables for the Cognitive, Emotional and Physical Clinical scales of the PPQ based on the regression models presented in Table 4, stratified by the significant sociodemographic predictors. The tables present raw scores that correspond to percentiles 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 95 (and associated *T* scores), with values calculated using the multiple regression equations derived from the healthy control group. Raw scores rounded to the nearest integer, rather than square root transformed scores on which the calculations were based, are presented for the ease of use. Where an individual's age falls between the stated values, then we recommend the individual's age should be rounded to the nearest value. Further, in cases where an individual's score is borderline abnormal (e.g., *T* score between 62.8 and 66.5), then calculation using the equations derived the regression models (Table 4) can be made to establish the precise *T* score (Van der Elst et al., 2006).

The final logistic regression models for the Validity scales (at a cut-off score of ≥ 5) are shown in Table 5. Multicollinearity was within acceptable limits (none of the independent variables had a standard error larger than 1.0) and there was no serious influence of outliers (maximum Cook's distance for the final models equaled .50). There was no gender effect on Validity scale scores in any domain. Age had a significant curvilinear effect on Validity scale scores in cognitive and emotional domains, reflecting a greater likelihood for participants in the youngest (18–29 years) and oldest age groups (60–69 years) to evidence scores ≥ 5 . Low education levels were linked with high scores on all Validity scales; a cut-off score of 5 yielded 29.0%, 19.4%, and 25.8% of participants as “likely non-credible” on Cognitive, Emotional, and Physical Validity scales, respectively, indicating poor specificity in this group. Specificity improved to 87.1%, 87.1%, and 90.3% on Cognitive, Emotional, and Physical Validity scales, respectively, when a cut-off score of 8 was adopted. The impact of age on scale scores within this group is unclear due to small numbers, although it was notable that of the 11 low education participants in the oldest age group (60–69 years), 5 (45.4%), 3 (27.3%), and 4 (36.4%) scored ≥ 5 on Cognitive, Emotional, and Physical Validity scales, respectively, suggestive of a potentially high rate of false positives in this group. Cumulative percentages for the observed values of the raw Validity scale scores as a function of significant predictors in logistic regression models are provided in Supplementary Table 4 (see Supplementary material online).

Comparison of Norms with ABI Participants

The mean Clinical index and subtest scores and Validity scale scores for ABI participants are reported in Table 6. As with the Community group, Clinical scales and subtests and Validity scales all showed adequate internal consistency of comprising items ($\alpha > .70$). ABI participants showed acceptable response consistency; the mean score on the response consistency index was 3.53 ($SD = 2.27$) with 75% of scores between 0 and 5. Further, no ABI participant had an inconsistency *T* score of more than 80, indicating an absence of excessive variable response inconsistency in the group.

Table 5. Final multiple logistic regression models for the Validity scale cut-off score ≥ 5 ($n = 403$)

| Validity scale total ≥ 5 | <i>B</i> | SE <i>B</i> | Wald χ^2 | OR (95% CI) | <i>R</i> ² |
|-------------------------------|----------|-------------|---------------|------------------------|-----------------------|
| Cognitive | | | | | |
| (constant) | -3.160 | 0.463 | 46.621*** | | |
| Age | -0.011 | 0.013 | 0.739 | 0.989 (0.964,1.014) | |
| Age ² | 0.002 | 0.001 | 7.494** | 1.002 (1.001,1.004) | |
| Education—Low | 1.669 | 0.565 | 8.735** | 5.307 (1.754,16.050) | |
| Education—High | -0.117 | 0.439 | 0.070 | 0.791 (0.376,2.105) | 0.114 |
| Emotional | | | | | |
| (constant) | -3.385 | 0.490 | 47.679*** | | |
| Age | -0.019 | 0.012 | 2.423 | 0.981 (0.957,1.005) | |
| Age ² | 0.002 | 0.001 | 6.889** | 1.002 (1.001,1.004) | |
| Education—Low | 1.481 | 0.630 | 5.526* | 4.399 (1.279,15.130) | |
| Education—High | 0.534 | 0.446 | 1.435 | 1.706 (0.712,4.086) | 0.063 |
| Physical | | | | | |
| (constant) | -4.094 | 0.713 | 32.978*** | | |
| Education—Low | 3.038 | 0.823 | 13.639*** | 20.870 (4.161,104.661) | |
| Education—High | 1.652 | 0.750 | 4.850* | 5.217 (1.199,22.696) | 0.108 |

Data regarding education was missing for one female participant; hence the model includes 403 participants only; Coding of the predictors: age = calendar age - 40.72; age² = (calendar age - 40.72)²; Education: Education—Low = 1; average or high education = 0; Education—High: high education = 1; average or low education = 0. SE = standard error; OR = odd ratios, CI = confidence intervals; *R*² = Nagelkerke *R*².

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 6. Summary statistics for ABI participants on the PPQ Clinical and Validity scales ($n = 59$)

| Clinical domain | Subscale (0–20) | Raw score mean (SD) | Range | α (95% CI) | T score mean (SD) |
|------------------------------|---------------------------|---------------------|-------|-------------------|---------------------|
| Cognitive | Language & Communication | 6.34 (3.80) | 0–16 | 0.75 (0.65,0.84) | |
| | Memory | 7.56 (4.79) | 1–19 | 0.86 (0.80,0.91) | |
| | Concentration | 7.66 (4.70) | 0–18 | 0.88 (0.82,0.92) | |
| | Executive | 6.02 (4.26) | 0–18 | 0.83 (0.75,0.89) | |
| Emotion | Anger | 5.31 (4.65) | 0–17 | 0.88 (0.83,0.92) | |
| | Anxiety | 4.86 (3.92) | 0–16 | 0.82 (0.75,0.88) | |
| | Depression | 5.73 (4.25) | 0–18 | 0.83 (0.75,0.89) | |
| | Stress | 5.27 (4.32) | 0–18 | 0.86 (0.80,0.91) | |
| Physical | Pain | 5.17 (4.01) | 0–18 | 0.82 (0.74,0.88) | |
| | Neurological | 5.49 (2.89) | 1–14 | 0.53 (0.32,0.69) | |
| | Disability & Restrictions | 6.10 (4.08) | 0–17 | 0.75 (0.64,0.83) | |
| | Somatic | 4.05 (3.19) | 0–15 | 0.72 (0.60,0.81) | |
| Scale totals (0–80) | | | | | |
| | Cognitive | 27.58 (15.22) | 4–68 | 0.94 (0.92,0.96) | 61.43 (10.60) |
| | Emotional | 21.17 (15.68) | 0–64 | 0.95 (0.94,0.97) | 56.39 (12.37) |
| | Physical | 20.81 (11.61) | 1–60 | 0.89 (0.85,0.93) | 64.31 (11.59) |
| Validity scale totals (0–24) | | | | | |
| | Cognitive | 2.69 (3.05) | 0–19 | 0.76 (0.66,0.84) | |
| | Emotional | 2.34 (3.45) | 0–21 | 0.82 (0.74,0.88) | |
| | Physical | 2.02 (3.08) | 0–20 | 0.79 (0.70,0.86) | |

In the few cases (<1%) where a participant failed to respond to one item on a subscale, the total subscale score was calculated by imputing the mean item score for the missing item; α = Cronbach's alphas; CI = confidence interval (calculated using Feldt's (1965) formula); Cronbach's alphas were calculated using data from participants who provided responses to all items on the relevant subscale.

Table 7. Cumulative percentage of ABI participants scoring at or selected raw score cut-offs on the Validity scales of the PPQ ($n = 59$)

| Validity scale totals | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 12 | 14 |
|-----------------------|------|------|------|------|------|------|------|------|------|------|------|
| Cognitive (0–24) | 22.0 | 39.0 | 54.2 | 71.2 | 86.4 | 91.5 | 94.9 | 94.9 | 98.3 | 98.3 | 98.3 |
| Emotional (0–24) | 42.4 | 55.9 | 66.1 | 72.9 | 79.7 | 81.4 | 93.2 | 98.3 | 98.3 | 98.3 | 98.3 |
| Physical (0–24) | 37.3 | 54.2 | 72.9 | 83.1 | 88.1 | 89.8 | 96.6 | 98.3 | 98.3 | 98.3 | 98.3 |

Raw scores on Clinical subscales were, generally speaking, positively skewed, although all were numerically greater than those of Community participants (Table 2) with wider ranges. The mean T scores for the Clinical scales, calculated using the regression-based norming procedure described earlier, indicated higher T scores (than norms) on all scales ($p < .001$; the assumption of homogeneity of variance was not violated in any comparison—for all comparisons, $p > .142$). The proportions of ABI participants exhibiting high numbers of problems (T score ≥ 65) on Cognitive (39.7%), Emotional (19.0%) and Physical Clinical scales (50.0%) were markedly greater than Community participants (6.9%, 5.2%, and 7.2%, respectively; for all group comparisons $p < .001$). Within-group comparisons of T scores indicated more severe problems in the cognitive and physical domains rather than with emotions ($p < .001$).

Mean Validity scale totals were low in the ABI group, with the majority of scores on each scale falling in the 0–2 range (Table 7). At the cut-off score of 5, specificity was 85% and 90% on the Cognitive Validity and Physical Validity scales, respectively, and slightly less than 80% on the Emotional Validity scale. After adjusting for significant demographic predictors (control group), there was a significant difference in the proportion of ABI participants classified as “likely non-credible” compared with the control group on the Emotional Validity scale ($p = .005$, odds ratio = 2.931, 95% CI = 1.383–6.211), but not on Cognitive Validity ($p = .219$) or Physical Validity scales ($p = .183$).

Discussion

The present study sought to develop a normative database that will allow clinicians and researchers to use the PPQ with patients with disabilities and cognitive impairments and assess their complaints, as well as the validity of their self-report. The data from the community sample provides the basis for interpreting the various PPQ scales, including ascertaining the demographic factors that potentially influence responding, and allows clinicians to determine whether scores are elevated or within normal limits relative to community-dwelling, non-disabled individuals. The community sample endorsed few items on the Clinical Scales, this being in line with what would be expected given that the sample were individuals with no neurological

or psychiatric conditions and none were undergoing treatment or attending clinical services. In contrast, the ABI group obtained elevated T scores on the Total Clinical scales, this being consistent with expectation. Using a cut-off of $T \geq 65$ to define a significant elevation, then between 19% and 50% of the ABI group exceeded that level on the Total Clinical scales indicating that they had significant difficulties in those areas. It is to be expected that a substantial proportion do not exceed the cut-offs bearing in mind the heterogeneous nature of the clinical sample. In both groups, response consistency was high with no case of excessive variable response inconsistency in the ABI sample. This confirms that the Clinical scales are sensitive to the cognitive, emotional, and physical issues with which ABI patients present and that the questionnaire is likely to be useful in a clinical context, such as an outpatient or rehabilitation setting. The satisfactory response consistency in the ABI group also indicates that responding to the questionnaire is not adversely affected by cognitive difficulties, an important consideration when assessing over-reporting in a brain injury population.

The factors influencing responding on the Total Clinical scales were multifactorial with gender, age, and educational level each contributing, and regression equations were developed for predicting T scores in each of the three domains which take into account their influence in contributing to self-report on the questionnaire. The relevance of sociodemographic factors for PPQ Clinical scale scores is not surprising. It is well established that poor self-reported health is more prevalent in individuals with low levels of education (Borgonovi & Pokropek, 2016; Schutte, Chastang, Parent-Thirion, Vermeulen, & Niedhammer, 2013), with evidence that less educated respondents tend to endorse more symptoms on questionnaires (Ladwig, Marten-Mittag, Formanek, & Dammann, 2000; Merten et al., 2016). The observed curvilinear effect of age in cognitive and emotional clinical domains is also supported by literature suggesting that middle-aged adults report higher levels of emotional, social, and psychological well-being than younger and older adults (Carstensen, Pasupathi, Mayr, & Nesselroade, 2000; Webster, Westerhof, & Bohlmeijer, 2012).

As expected, the community sample had uniformly low scores on the Validity scales, as did the majority of the ABI group whose scores on each fell predominantly in the 0–2 range. When appropriate adjustments were made for the demographic predictors, there was no significant difference between the Cognitive and Physical Validity scales for the ABI and Community groups, although the ABI group differed on the Emotional Validity scale. However, in practical terms, this was a minor difference, the equivalent of endorsing “sometimes” on only one of the validity items, and readily addressed by raising the cut-off to ≥ 6 . These findings are reassuring and indicate that despite having had a neurological injury, and potentially having cognitive impairments, the ABI participants did not endorse having non-valid symptoms, which confirms the utility of the questionnaire.

Generally speaking, SVTs should be able to detect non-credible symptom reporting while being relatively insensitive to sociodemographic indicators. Interestingly, across all three Validity scale domains, control participants with low education were more likely to endorse items. The results of the consistency index suggest that it is unlikely these participants simply failed to understand items. However, it is worth considering that education level has been (inversely) associated with over-reporting of symptoms on the Structured Inventory of Malingered Symptomatology (Dandachi-Fitzgerald, 2017) and with “pseudosymptom” endorsement on the recently developed SRSI (Merten et al., 2016). It is possible less educated people are more likely to feign or overreport symptoms in a more obvious manner than those with higher levels of education, who may tend toward more subtle methods of feigning or over-reporting (Dandachi-Fitzgerald, 2017; Solomon, Boone, Miora, & Skidmore, 2010). Assessing symptom validity in participants with low education using dedicated scales may, therefore, present a unique challenge, with an elevated risk of false-positive classifications unless appropriate threshold adjustments are made. The interpretation of the curvilinear relationship of age with performance validity is unclear. Age effects have been observed on PVTs such as the WMT in patients with traumatic brain injury (Sherer et al., 2015), although not often (Green, 2003). Clearly, future work is needed to address age differences in Validity scale responses and their relationship with education level.

Nevertheless, interpretation of the PPQ can follow a clear course with the clinician initially examining the Consistency measure to confirm that the questionnaire has been responded to in a reliable manner. Unreliable responding (i.e., Consistency T score > 80) would indicate that, for whatever reason, the examinee has not responded to the PPQ consistently, and further interpretation should not be pursued. Assuming there are no concerns about the reliability of responding, then the three Validity scales should be examined to determine whether one or more are elevated, the most straightforward method being to examine the cumulative percentages of scores in Table 3 and deciding on a level of specificity suited to the clinicians’ particular setting and requirements. A cut-off score of ≥ 5 provides satisfactory specificity, that is $\geq 90\%$, for the Cognitive and Physical Validity scales. Elevated scores beyond this cut-off call into question the validity of the patient’s self-report and ≥ 6 provides excellent specificity for both the community and patient sample, with significant elevations (e.g., ≥ 8) being particularly unlikely and non-credible. Clinicians can also take into account the patient’s educational level, and for those with a low level of education, a cut-off of ≥ 8 would be appropriate, although given the limited number of those with low education, interpretation should be undertaken cautiously. Elevations on one Validity scale, but not others, may indicate that the respondent is responding selectively in a non-credible manner and call into question their reporting in that area, but potentially

not in others, and the associated Clinical scales might be treated with circumspection. In such circumstances, elevated Clinical scales likely reflect over-reporting, rather than true impairment. On the other hand, when all the Validity scales are within normal limits, the Total Clinical scales are each interpretable. For example, a 20-year-old man with brain trauma and a high educational background who obtained a Total Cognitive score of 38, Emotional score of 23 and Physical score of 6, would have corresponding *T* scores of 62.8 (90th percentile), 55.2 (70th percentile), and 47.5 (40th percentile), respectively, indicating that his difficulties were predominantly cognitive and emotional and he had little in the way of physical disability (see Supplementary material online, *Tables 1–3*). In the event that a Total Clinical scale is elevated, the four individual subscales can be examined to determine the source of the elevation and the particular area of difficulty. A unique advantage of the PPQ compared with other multidimensional measures is that elevations on the validity scales represent prima facie evidence of non-credible responding, whereas over-reporting on other measures, such as the PAI, may be consistent with the severity of the individual's disorder, as typically they involve a frequency count of genuine symptoms. In the case of the PPQ, where the validity symptoms are a priori implausible, then their endorsement is inherently inconsistent with a valid condition.

Limitations and Future Studies

The PPQ is shorter than other multidimensional measures, whereas nevertheless allowing a broad assessment of patients' complaints and including measures that assess response validity, and the present study provides potentially useful data to assist with its use and interpretation. Clearly, research on the PPQ is in its early phase with more work is needed. The ABI sample was a heterogeneous group with a range of neurological conditions which may be useful as an omnibus clinical sample, but future investigation might look at a more uniform group such as TBI patients in both clinical and forensic settings. In addition, concomitant neuropsychological test data and data from an established personality scale which contains emotional, cognitive and physical item scales and validity scales for each of these areas (e.g., MMPI-2-RF) are needed to allow examination of the relationship of PPQ scales to established tests of actual ability/impairment and personality. Further research is also required to examine its utility in assessing over-reporting in other populations, such as those with chronic pain and illness behavior, such as somatisation and non-epileptic attack disorder. The present study involved a relatively small clinical sample of whom a subset had traumatic brain injuries, a group that is of particular interest in medico-legal settings. [van den Broek and colleagues \(2012\)](#) found that the validity scales had good sensitivity and specificity when discriminating between simulated brain trauma, patients with acquired brain injuries and controls. The present study sought to establish that the validity measures were unaffected by neuropathology in a non-forensic sample, and this was largely confirmed, but additional investigation is required to evaluate the utility of the scales with real-life TBI patients involved in personal injury litigation.

Supplementary Material

Supplementary material is available at *Archives of Clinical Neuropsychology* online.

Funding

None.

Conflict of interest

None declared.

Acknowledgements

This work was (managerially) supported by and undertaken at St. George's Healthcare NHS Trust, London.

References

- American Academy of Clinical Neuropsychology Board of Directors. (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, 21, 209–231.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: American Psychiatric Association.

- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York, NY: Wiley.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF. Minnesota Multiphasic Personality Inventory – 2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Bodapati, A. S., Combs, H. L., Pastorek, N. J., Miller, B., Troyanskaya, M., Romesser, J., et al. (2018). Detection of symptom over-reporting on the Neurobehavioral Symptom Inventory in OEF/OIF/OND veterans with history of mild TBI. *The Clinical Neuropsychologist*, 5, 1–18.
- Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, 23, 729–741.
- Borgonovi, F., & Pokropek, A. (2016). Education and self-reported health: Evidence from 23 countries on the role of years of schooling, cognitive skills and social capital. *PLoS One*, 11, e0149716.
- Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology*, 20, 419–426.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the Re-Standardised Minnesota Multiphasic Personality Inventory: MMPI-2*. Minneapolis: University of Minnesota Press.
- Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*, 79, 644–655.
- Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, 21, 559–566.
- Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. London: Chapman and Hall.
- Cragar, D. E., Bery, D. T., Fakhoury, T. A., Cibula, J. E., & Schmitt, F. A. (2006). Performance of patients with epilepsy or psychogenic non-epileptic seizures on four measures of effort. *Clinical Neuropsychologist*, 20, 552–566.
- Dandachi-Fitzgerald, B. (2017). *Symptom validity in clinical assessments* (Doctoral dissertation, Maastricht University).
- Etherton, J. L., Bianchini, K. J., Ciota, M. A., Heinly, M. T., & Greve, K. W. (2006). Pain, malingering, and the WAIS-III Working Memory Index. *Spine Journal*, 6, 61–71.
- Feldt, L. S. (1965). The approximate sampling distribution of Kuder–Richardson Reliability Coefficient Testing. *Psychometrika*, 30, 357–370.
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.
- Green, P. (2003). *Word Memory Test*. Edmonton, Alberta, Canada: Green's Publishing.
- Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., et al. (2006). The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *Neurotoxicology*, 27, 940–950.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R., Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129.
- Holdnack, J. A., Whipple Drozdick, L., Weiss, L. G., & Iverson, G. L. (2013). *WAIS-IV, WMS-IV, and ACS: Advanced Clinical Interpretation*. San Diego: Academic Press.
- Ladwig, K. H., Marten-Mittag, B., Formanek, B., & Dammann, G. (2000). Gender differences of symptom reporting and medical health care utilization in the German population. *European Journal of Epidemiology*, 16, 511–518.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410–425.
- Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.
- McMillan, T. M., Anderson, S., Baker, G., Berger, M., Powell, G. E., & Knight, R. (2009). Assessment of effort in clinical testing of cognitive functioning for adults. *Professional Practice Board-British Psychological Society*, 1–27.
- Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law*, 9, 102–111.
- Meyers, J. E., English, J., Miller, R. M., & Lee, A. J. (2015). Normative data for the Neurobehavioral Symptom Inventory. *Applied Neuropsychology: Adult*, 22, 427–434.
- Morey, L. C. (1991). *Personality Assessment Inventory Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Morgan, J. E., & Sweet, J. J. (2009). *Neuropsychology of malingering casebook*. New York: Psychology Press.
- Oosterhuis, H. E., van der Ark, L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, 23, 191–202.
- Parmenter, B. A., Testa, S. M., Schretlen, D. J., Weinstock-Guttman, B., & Benedict, R. H. (2010). The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, 16, 6–16.
- Roelofs, J., van Breukelen, G., Sluiter, J., Frings-Dresen, M. H., Goossens, M., Thibault, P., et al. (2011). Norming of the Tampa scale for kinesiophobia across pain diagnoses and various countries. *Pain*, 152, 1090–1095.
- Rogers, R. (2008). Detection strategies for Malingering and Defensiveness. In Rogers R. (Ed.), *Clinical Assessment of Malingering and Deception* (3rd ed., pp. 14–35). New York: The Guilford Press.
- Schütte, S., Chastang, J. F., Parent-Thirion, A., Vermeylen, G., & Niedhammer, I. (2013). Social differences in self-reported health among men and women in 31 countries in Europe. *Scandinavian Journal of Public Health*, 41, 51–57.
- Sherer, M., Davis, L. C., Sander, A. M., Nick, T. G., Luo, C., Pastorek, N., et al. (2015). Factors associated with Word Memory Test performance in persons with medically documented traumatic brain injury. *The Clinical Neuropsychologist*, 29, 522–541.
- Solomon, R. E., Boone, K. B., Miora, D., & Skidmore, S. (2010). Use of the Picture Completion Subtest as an embedded measure of effort; Abstracts of Division 40 Annual Meeting. *The Clinical Neuropsychologist*, 24, 936.
- Tombaugh, T. N. (1996). *TOMM: Test of Memory Malingering*. New York: Multi-Health Systems.
- Van Breukelen, G. J., & Vlaeyen, J. W. (2005). Norming clinical questionnaires with multiple regression: The Pain Cognition List. *Psychological Assessment*, 17, 336–344.

- van den Broek, M. D., Monaci, L., & Smith, J. G. (2012). Clinical utility of the Personal Problems Questionnaire (PPQ) in the assessment of non-credible complaints. *Journal of Experimental Psychopathology*, *3*, 825–834.
- van den Broek, M. D., & Sembi, S. (2017). Malingering and Exaggeration. In van den Broek M. D., & Sembi S. (Eds.), *Brain Injury Claims*. London: Sweet & Maxwell.
- Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2006). The Concept Shifting Test: Adult normative data. *Psychological Assessment*, *18*, 424–432.
- Webster, J. D., Westerhof, G. J., & Bohlmeijer, E. T. (2012). Wisdom and mental health across the lifespan. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *69*, 209–218.
- Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court*. New York: Springer.