Validity of symptom reports of asylum seekers in a psychiatric hospital: A descriptive study

Douwe van der Heide [a,*], Harald Merckelbach [b]

[a] Phoenix, ProPersona, Wolfheze, The Netherlands
[b] Forensic Psychology Section, Maastricht University, Maastricht, The Netherlands

ARTICLE INFO

ABSTRACT

Our study involved three samples ($N = 85$; $N = 38$, and $N = 27$) of asylum seekers in a Dutch psychiatric hospital. We looked at how often they reported severe dissociative episodes (i.e., not recognizing oneself in a mirror; seeing traumatic images in a mirror) and whether these symptoms were related to deviant performance on Symptom Validity Tests (SVTs), notably items from the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) and a forced-choice task modeled after the Morel Emotional Numbing Test (MENT; Morel, 1998). We also examined whether poor language proficiency and the presence of incentives to exaggerate symptoms might affect scores on SVTs. Dissociative target symptoms were reported by considerable percentages of patients (27–63%). Patients who reported these symptoms had significantly more often deviant scores on SVT items compared with those who did not report such symptoms. With a few exceptions, deviant scores on SVT items were associated with incentives rather than poor language skills. We conclude that the validity of self-reported symptoms in this target group should not be taken for granted and that SVTs may yield important information.

1. Introduction

Differences in culture and language complicate diagnostic decision making in psychiatry and may therefore affect its outcome negatively (Gara et al., 2012). Some authors have argued that it is below professional standards when clinicians evaluate patients whose language they do not speak (Artiola i Fortuny & Mullaney, 1998). While this may be true in a general sense, clinicians confronted with asylum applicants who present with urgent mental problems often have to rely on on-the-spot translations by professional interpreters (Lustig, 2008; Reko, Bech, Wohlert, Noerregaard, & Csillag, 2015).

Apart from cultural and language factors, intentional over-reporting may distort symptom reports and ultimately diagnostic judgment (Meffert, Musalo, McNiel, & Binder, 2010). That is, asylum applicants in countries such as the Netherlands may have positive incentives to exaggerate their mental health problems; if their case for political asylum has been rejected, a refugee status may still be granted for medical reasons. Even after a refugee status has been granted, the ex-asylum applicants may claim special benefits (e.g., housing and family reunion) for people with mental vulnerabilities. Thus, some asylum applicants may have a motive to over-report symptoms. The other side of the coin is that reporting psychiatric symptoms during asylum-seeking procedures might be associated with negative incentives (e.g., compulsory hospital admission and accordingly limited freedom of movement). Meanwhile, little is known about the prevalence of symptom over-reporting among asylum applicants who present with psychiatric problems. Some authors have argued that malingering among asylum applicants is rare (Lustig, 2008), whereas others opine that malingering is a distinct possibility that should be taken into account (Morgan, 2007). Curiously enough, in the extant literature on, for example, post-traumatic stress symptoms in asylum seekers there is no example of a study in which researchers tried to differentiate between genuine symptom reports and symptom over-reporting. A case in point is the study of Kissane, Szymanski, Upthegrove, and Katona (2014) who conducted interviews to assess post-traumatic stress symptoms in asylum seekers who reported psychological problems. More than half of the patients apparently suffered from severe post-traumatic symptomatology, and yet this statistic is difficult to interpret because the authors did not rule out symptom exaggeration. While Kissane et al. (2014) do acknowledge the cross-cultural validity problems that may arise when administering an interview to patients originating from Africa and Asia, they do not mention symptom exaggeration as a source of potential bias.

One way to assess the validity of symptom reports is by administering items that check for a tendency to claim implausible symptoms and deficits. There is now an extensive literature on so-called Symptom Validity Tests (SVTs) in the forensic context (see for a review

* Corresponding author at: GGZ Centraal, P.O. Box 1000, 3850 BA Ermelo, The Netherlands.
  E-mail addresses: douwevanderheide@gmail.com (D. van der Heide),
h.merckelbach@maastrichtuniversity.nl (H. Merckelbach).

e.g., Sollman & Berry, 2011), but little research has been done on their utility in a cross-cultural context such as evaluating the symptoms of asylum applicants. However, the few cross-cultural studies that have been conducted suggest that SVTs may contribute to diagnostic accuracy in this domain (Cirlugea, 2014; Montes & Guyton, 2014; Salazar, Lu, Wen, & Boone, 2007). Still, so far, no study has made an attempt to differentiate between language difficulties and intentional over-reporting as contributory factors to poor validity of symptoms reported by asylum applicants (see also Drob, Meehan, & Waxman, 2009).

The current study explored the utility of typical SVT items in the diagnostic evaluation of asylum applicants who had been admitted to a psychiatric hospital. In this paper, asylum applicants may refer to refugees as well. In the Dutch system, refugees are asylum applicants who are granted a refugee status. In the first sample, we conducted with the help of professional interpreters an open interview, to examine how often asylum applicants admitted to a psychiatric hospital endorse specific depersonalization symptoms. We focused on symptoms that many clinicians assume to reflect a history of trauma (Friedman et al., 2011). In samples 2 and 3, we focused on similar symptoms and explored whether they are related to over-endorsement of implausible symptoms. We also tested whether limited proficiency in the language of the host country (Dutch) may be responsible for such over-endorsement.

Our study was exploratory in nature and the two main questions that we wanted to address were as follows: are severe dissociative symptoms reported by asylum seekers in a psychiatric hospital associated with deviant scores on SVT items? And to what extent does deviant SVT performance reflect poor language skills or incentives to exaggerate?

2. Method

2.1. Patients

Sample 1 consisted of a consecutive case series of 85 asylum seekers (60 men), who had been admitted to a clinical facility specialized in the treatment of asylum applicants with severe mental disorders in the Netherlands. This facility offers non-forensic, inpatient treatment only (32 beds) and is part of a larger general psychiatric hospital. Mean age was 37.3 years ($SD = 10.9$; range 18–61 years). A total of 28 patients (33%) originated from the Middle East, 23 (27%) from the former USSR, 23 (27%) from Africa, seven (8%) from former Yugoslavia, and four (5%) from East Asia. A majority (67%) was reported to have a history of past psychotic episodes, and 27 patients (32%) were still considered psychotic upon referral to the facility.

Sample 2 involved a consecutive series of 38 asylum seekers (30 men), who were treated in the same facility, after excluding three patients who were too disorganized to undergo testing and another two who refused informed consent. Mean age in this sample was 35.2 years ($SD = 11.7$; range 18–61 years). In this sample, six patients (16%) came from the Middle East, 12 (32%) from the former USSR, 16 (42%) from Africa, three (8%) from former Yugoslavia, and one (3%) from East Asia. Eight patients (21%) had a poor proficiency in Dutch, 14 (37%) had obtained an intermediate level of proficiency in that language, and 16 (42%) were advanced students, as determined by a set of widely used criteria (see below). Also, 20 patients (53%) had a positive incentive to over-report symptoms, 13 (34.2%) had no incentive, and five (13.2%) had a negative incentive as judged by an independent review of their case files (see below).

Sample 3 consisted of a consecutive series of 27 asylum seekers (19 men) from the same facility, after three patients were excluded because of severe mental disorganization and another two because they refused informed consent. Their mean age was 34.5 years ($SD = 9.7$; range 20–51 years). In this sample, eight patients (30%) originated from the Middle East, three (11%) from the former USSR, 11 (41%) originated from Africa, one (4%) from former Yugoslavia, and four (15%) from

East Asia. Ten patients (37%) had a poor proficiency in Dutch, nine (33%) an intermediate proficiency, and eight (30%) a good proficiency. Furthermore, 12 patients (44%) had a positive incentive to over-report symptoms, nine (33%) no incentive, and six (22%) a negative incentive.

The study was conducted between 2005 and 2009. All patients admitted to the facility were included, except for those who refused informed consent or who were too disorganized to give such consent. The first sample was recruited two years before data in the second and the third sample were collected. Thus, there was no overlap in patients between sample 1 and samples 2 and 3. The time interval between data collection in samples 2 and 3 was several weeks. Thus, when recruitment of sample 3 began, eight patients from sample 2 were still admitted in the clinic. Accordingly, they were included in sample 3 as well. The combined samples 2 and 3 consisted of 57 asylum seekers (45 men) with a mean age of 35.4 years ($SD = 10.9$; range 18–61 years). Of these 13 patients (23%) originated from the Middle East, 13 (23%) from the former USSR, 24 (42%) from Africa, three (5%) from former Yugoslavia, and four (7%) from East Asia. Sixteen patients (28%) had a poor proficiency in Dutch, 19 (33%) an intermediate proficiency, and 22 (39%) a good proficiency. Furthermore, 28 (49%) had a positive incentive to over-report symptoms, 22 (39%) no incentive, and seven (12%) a negative incentive.

2.2. Measures

2.2.1. Open interview (sample 1)

The purpose of the open interview was to explore on which scale patients reported the so-called *signe du miroir* (i.e., the inability to recognize oneself in a mirror) and related depersonalization symptoms. Ever since the French psychiatrist Paul Abély (1927) described the *signe du miroir*, it has been regarded as a severe form of depersonalization. Some authors concede that this symptom often precedes a psychotic breakdown (Abély, 1927; Goedhart & Sno, 2014) and is related to a traumatic history (Friedman et al., 2011). Although the *signe du miroir* was first described by French psychiatrists, it has been reported for non-Western psychiatric samples as well (Yu-Fen & Neng, 1981). This symptom had the special attention of psychiatrists involved in the care for asylum applicants in the facility, because psychotic episodes require additional treatment effort. With these considerations in mind, the first author (DvdH) asked asylum applicants in the context of routine psychiatric evaluations about any particular experiences with mirrors during their admission. Questions were ad hoc translated by professional interpreters, who were either present in person or provided their services over the phone.

2.2.2. Implausible symptoms (samples 2 and 3)

We used items derived from the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) to assay the tendency to exaggerate symptoms. In its original form, the SIMS is a self-report instrument designed to screen for exaggeration of neurocognitive and psychiatric complaints. Basically, it consists of 75 true-false items that describe atypical and rare symptoms and experiences (e.g., "There is a constant ringing in my ear"; "The voices that I hear, have never stopped since they began"). A clear advantage of SIMS items is that they are easy to understand. There are five subscales, each containing 15 items, which address commonly feigned conditions: amnesia, neurologic impairment, psychosis, affective disorders, and low intelligence. After recoding some items, endorsed symptoms are summed to obtain a total SIMS score, with higher scores indicating more symptom over-endorsement. Previous studies recommended a cutoff of 16 for a comprehensive assessment of feigning (Merckelbach & Smith, 2003). Van Impelen, Merckelbach, Jelicic, and Merten (2014) summarize psychometric data indicating that the internal consistency of the SIMS is satisfactory (with Cronbach's alpha's ranging from 0.80 to 0.96), its test–retest stability sufficient (r's = 0.72–0.97), and its ability to

discriminate between symptom exaggeration and honest responding fairly effective (with sensitivities varying between 0.75% and 100%).

For the purpose of the current study, some items of the Dutch research version of the SIMS (Merckelbach, Koeyvoets, Cima, & Nijman, 2001) were adapted. For example, after consultation with certified translators, the item pertaining to the queen of Holland was rephrased as follows: "The prophet of Allah is called Mohammed" (for patients with a Muslim background) and "The mother of Jesus is called Mary" (for patients with a Christian background).

For sample 2, two items about symptoms that had been found to be important during the open interview (sample 1; see above) were embedded as additional items in the list of SIMS items. These additional items were: "Some people have the experience of looking in a mirror and not recognizing themselves. Does this happen to you?" and "When I look at myself in the mirror, I think about terrible things that happened in my past". After instructions had been given and had been translated, the SIMS symptoms and the additional items were read out aloud. This was done by a fourth year psychology student for sample 2 and by the first author for sample 3. Items were presented at a comfortable pace, and each item was translated by professional interpreters, who also interpreted the answers of the patients (either a yes or a no).

2.2.3. Forced-choice task (samples 2 and 3)

Patients were administered a forced-choice task closely modeled after the Morel Emotional Numbing Test (MENT; Morel, 1998). The task aims to detect response distortion in the assessment of trauma-related problems. We used a version that was developed by Geraerts et al. (2009), who observed in their sample of Croation war veterans that high error levels were highly effective in differentiating between treatment-seeking and compensation seeking veterans (sensitivity: 92%; specificity: 96%). Briefly, the task comprised 20 colored slides of ten facial expressions posed by a man and a woman. Their expressions reflected happiness, frustration, sadness, anger, fear, calmness, surprise, shyness, confusion, and sleepiness. The slides were presented on a computer screen along with simple words that described emotional expressions (e.g., "happy"; "angry"). In a first series of 20 trials, patients saw one expression on the computer screen and had to indicate which of two words (e.g., "happy" versus "surprised") described the facial expression. In a second run of 20 trials, patients viewed two slides of different expressions, and they were given only one word; their task was to identify the expression that best matched the word. In a final run of 20 trials, patients were shown two slides and were given two words; slides and words had to be matched. The tests were conducted by the first author. During the test procedure, a professional interpreter was present and assisted with translating the instructions and the key verbal labels.

Before they underwent the task, patients were told that emotional numbness is a prominent symptom of trauma-related problems and that this may cause people to have difficulties with the recognition of facial expressions. The rationale behind this instruction is that individuals who want to overstate trauma-related symptoms may intentionally produce more errors during the forced-choice procedure. Errors were summed across the three runs. Morel (1998) recommended a cutting score of nine errors, with scores above this level raising the suspicion of symptom over-reporting. In the current study, we adopted the cutting score of nine errors (see also Geraerts et al., 2009). Compared with the open interview or the symptoms of the SIMS, forced-choice tasks such as the MENT require only minimal effort to translate test items, which makes them interesting in cross-cultural contexts. Another reason to employ this type of forced-choice task is that it relies on the ability of respondents to identify facial expressions of basic emotions. Ever since Darwin (1872) and the pioneering work of Ekman and Friesen (1971) we know that expressions of basic emotions have their roots in biology rather than in culture.

2.2.4. Dissociative experiences (sample 3)

Patients were given the items of the Dissociative Experiences Scale (DES-II; Bernstein & Putnam, 1986). The DES is a self-report scale that requires participants to indicate on 100 mm visual analog scales (VAS; anchors: 0 = never; 100 = always) to what extent they experience 28 dissociative experiences in daily life (amnesia, depersonalization, absorption). In a meta-analysis, van IJzendoorn and Schuengel (1996) provide evidence for the sound psychometric properties of the DES. Summarizing the findings of a large pool of studies, among which studies that administered the DES to African American, Caucasian, and Hispanic war veterans, these authors conclude that the overall internal consistency of the DES is good (mean Cronbach's alpha = 0.93), while the test–retest stability is satisfactory (rs = 079–0.90). The DES items were read aloud by a sixth year medical resident, after which each item was translated by a professional translator, who also interpreted the answers of the patients. One of the two target symptoms that had been explored during the open interview in sample 1 – not recognizing yourself in a mirror – is listed as a separate item in the DES (i.e., item 11). Therefore, there was no need to add this item to the list of SIMS symptoms, as we did in sample 2. The second target symptom was added as an extra item to the DES and rephrased to fit the original description better, which specifically included seeing images ("Some people have the experience of looking in a mirror and seeing images of people or events from the past"). DES items as well as the additional item were rated by patients on 0–100% VAS (0 = 0% of the time, 100 = 100% of the time). The target symptoms – not recognizing yourself in a mirror and seeing images from the past in mirrors – were considered to be present when they were endorsed at a minimum of 10%.

Because the pertinent depersonalization symptoms were gauged in a different way in samples 2 and 3 – within the context of the SIMS items and within the context of the DES items, respectively – we decided to present the data collected in both samples separately. The supplemental file, however, contains the SIMS and forced-choice data collapsed across samples 2 and 3 (N = 57).

2.3. Procedure

The SIMS items, the forced-choice task, and DES items were administered to patients after the hospital decided to introduce Routine Outcome Monitoring (ROM). SIMS items and forced-choice task were included to serve as an additional quality check on self-reported symptoms. Before SIMS items and forced-choice task were administered, patients were informed that the tests were employed to assess the validity of symptom reporting in their target group. They were told that if the test results indicated a poor validity in their case, the conclusion would be that Western-style psychological tests would not provide useful information and that their diagnosis was to be based on additional interviews and observations. Only patients who gave informed consent for anonymous use of their data for scientific purposes were included. The study was approved by the Central Committee on Research Involving Human Subjects (CCMO). SIMS items, forced-choice task, and dissociative experiences items were presented in a counterbalanced fashion as much as possible.

Patients in samples 2 and 3 were independently assessed by Dutch language teachers of the hospital. On the basis of fixed criteria (Meijer & Noijons, 2008), these experts categorized patients into speakers with a poor proficiency in Dutch, an intermediate proficiency, and an advanced proficiency. These proficiency levels were taken to be a reasonable proxy for the degree to which patients would need the help of professional interpreters during diagnostic procedures.

For samples 2 and 3, social workers of the hospital independently evaluated asylum applicants' files for the presence of incentives. The social workers were blind as to patients' performance on the SIMS items and the forced-choice task. Three groups were formed: patients with a positive incentive to over-report symptoms, patients who had as many positive as negative incentives (referred to as "no" incentive

group), and patients with a negative incentive. Patients were assigned one point for each circumstance promoting over-reporting, notably: 1. involvement in an ongoing asylum procedure; 2. seeking a temporary refugee status issued for medical reasons; and/or 3. involvement in any other current procedure requiring a medical report indicating medical necessity, urgency or exemption (e.g., request for family reunion although the patient is not able to generate the necessary income demanded by Dutch law; a request for urgent change of housing or special housing arrangements; a request to be exempted from the demand to pass the language test in the naturalization procedure). For each circumstance that would make intentional over-reporting less likely, the raters subtracted a point. Such circumstances would be: 1. a compulsory nature of the present admission; 2. involvement in any current procedure requiring a medical report indicating improved functioning or decreased need for medical treatment or scrutiny (e.g., a child custody procedure, a request for voluntary repatriation). Patients with one point or more were considered to have a positive incentive, and patients with minus one point or less to have a negative incentive.

Importantly, neither the psychiatrist who conducted the open interviews, nor the psychiatrists or students who presented the SIMS items, forced-choice task, and DES items were aware of the language proficiency level of applicants. Neither had they had information regarding patients' incentive scores. On the other hand: both the psychiatrist and the students were involved in the treatment of the patients. So they had some background information about patients and blinding was not complete.

2.4. Data analysis

We used descriptive statistics to study the prevalence of severe dissociative target symptoms, SIMS symptoms, and errors on the forced-choice task in our samples. Depending on whether data were skewed or evenly distributed, we employed one-way Analyses of Variance (ANOVA's) or Kruskal–Wallis and Mann–Whitney U tests to compare language proficiency and incentive groups with regard to their endorsement of SIMS symptoms and their errors on our forced-choice task.

3. Results

3.1. Prevalence of target symptoms in sample 1

Thirty patients (35%) reported during the open interview the experience of not recognizing themselves in mirrors. That is, they regularly had *at least* doubts as to whether the reflection in the mirror was their own. Twenty three patients (27%) said that they had experienced seeing images in the mirror with a traumatic content, usually pertaining to aversive events from their personal past or related to loved ones who passed away during traumatic circumstances. Both symptoms had in common that they appeared to be induced by stress and that they had a tendency to occur during the night. Patients reporting these symptoms also said that they tended to avoid mirrors, except for five patients who said that they obsessively scrutinized their appearance in mirrors. The capacity to test the reality of these experiences remained intact in all patients, except for a minority who met criteria for a self-misidentification syndrome (four patients), or for a paranoid delusion towards mirrors (five patients). Thus, the results of the open interview showed that about one third of the patients reported the *signe du miroir* and related dissociative experiences.

3.2. Endorsed SIMS symptoms and forced-choice performance in sample 2

Eleven respondents said that they were not able to choose between "true" or "false" for a total of 71 SIMS symptoms. These missing values were treated in a conservative way, i.e., as indicating non-endorsement. The internal consistency of the SIMS symptoms was satisfactory. Thus, Cronbach's alpha was 0.95 for the total set of items and

ranged from 0.56 (affective disorders subscale) to 0.86 (psychotic disorders subscale) for separate subscales. As a check on inter-rater reliability of the SIMS items, eight patients were tested twice (with time intervals in between of several weeks): by the first author (DvdH) and by a test assistant. In this subgroup, test–retest scores correlated significantly: $r = 0.91$, $p < 0.05$.

The mean endorsement rate of SIMS symptoms was 35.1 ($SD = 15.7$), 95% CI [29.7, 40.5]. A majority of patients (87%) had scores that exceeded the original cutoff of 16. Patients exhibiting poor Dutch proficiency ($n = 8$) endorsed on average 44.1 SIMS symptoms ($SD = 8.8$); those with an intermediate proficiency ($n = 14$) obtained a mean score of 34.1 ($SD = 18.3$), and those with a good proficiency ($n = 16$) had a mean endorsement rate of 31.5 ($SD = 15.0$). The SIMS data in this sample were normally distributed. A one-way ANOVA that compared the mean SIMS scores of the three groups failed to attain significance: $F(2, 35) = 1.84$, $p = 0.17$. Patients with positive incentives ($n = 20$) had a mean SIMS score of 44.2 ($SD = 11.3$), those with a negative incentive ($n = 5$) had a mean score of 13.8 ($SD = 6.2$), while those with "no" incentives ($n = 13$) had an intermediate score of 29.3 ($SD = 13.6$); a one-way ANOVA yielded a significant effect, $F(2, 35) = 15.9$, $p < 0.001$, eta$^2 = 0.48$. Follow-up t-tests made it clear that the positive incentive group had a higher SIMS endorsement rate than the negative incentive group ($t (23) = 5.7$, $p < 0.05$) and the group with no incentives ($t (31) = 3.4$, $p < 0.05$); the negative incentive group had lower endorsement levels than the group without incentives ($t (16) = 2.4$, $p = 0.05$).

There were no missing data for the forced-choice trials in this sample. Cronbach's alpha was 0.96. The mean error score on our forced-choice task was 17.7, 95% CI [12.6, 22.8]. In total, 58% of the patients scored above the cut-point of nine errors. All patients who scored above the cut-off for the SIMS, also did so for the forced-choice task. The correlation between the number of forced-choice errors and endorsement rate of SIMS symptoms was $r = 0.65$, $p = 0.01$.

Patients with a poor proficiency had on average 32.3 errors ($SD = 13.1$), those with an intermediate proficiency had 17.0 errors ($SD = 11.7$), and patients with a good proficiency made 11.1 errors ($SD = 13.7$). The forced-choice data in this sample were not normally distributed (Shapiro–Wilk $p < 0.001$). A Kruskal–Wallis test showed that group differences were significant: $p < 0.01$. Patients with a positive incentive had 27.0 errors ($SD = 14.8$), those with a negative incentive had on average 5.6 errors ($SD = 3.5$), while patients with "no" incentive had 8.2 errors ($SD = 5.5$). A Kruskal–Wallis test indicated that these group differences were significant: $p < 0.001$.

The additional item on having difficulties recognizing oneself in a mirror was endorsed by 14 patients (37%). The item on thinking of terrible past events when looking into a mirror was endorsed by 24 patients (63%). Table 1 shows how these symptom reports relate to SIMS and forced-choice scores. As can be seen, patients who said that they had difficulties recognizing themselves in mirrors endorsed more SIMS symptoms, $t (38) = 4.17$, $p < 0.01$, and had more forced-choice errors, Mann–Whitney U: $p < 0.01$, compared with patients who did not report this symptom. A similar pattern emerged for reporting thoughts about terrible past events when looking into mirrors: $t (38) = 7.43$, $p < 0.01$ and Mann–Whitney U: $p < 0.01$, respectively.

We also looked at percentages of patients who failed the SIMS and forced-choice cut-points and basically replicated the pattern described above, but for one exception. Those who reported not being able to recognize themselves in mirrors more often failed on the forced-choice task, but not on the SIMS items: Fisher exact $ps = 0.02$ and 0.14, respectively. For thinking about terrible past events when looking into mirrors, both Fisher exact ps were <0.01.

3.3. Endorsed SIMS symptoms and forced-choice performance in sample 3

In this sample, the SIMS items were presented to patients by the first author; there were no missing data. Cronbach's alpha of the total set of

Table 1
Number of endorsed SIMS items and forced-choice errors of patients in sample 2 ($N = 38$) who did (yes) or did not (no) report the target symptoms of depersonalization.

	Not recognizing oneself in the mirror		Thinking about the terrible past events when looking into the mirror	
	Yes ($n = 14$)	No ($n = 24$)	Yes ($n = 24$)	No ($n = 14$)
Mean (SD) SIMS	46.7 (11.1)	28.3 (14.1)	44.3 (10.7)	19.3 (8.8)
n (%) > 16 SIMS	14 (100%)	19 (79%)	24 (100%)	9 (64%)
Mean (SD) forced-choice errors	29.1 (15.1)	11.1 (10.1)	24.0 (15.5)	7.1 (3.8)
n (%) > 9 errors Forced-choice	12 (86%)	10 (42%)	18 (75%)	4 (28%)

SIMS symptoms was 0.96 and alpha's for subscales varied between 0.74 (affective disorders) and 0.91 (amnesia). The mean SIMS symptom endorsement rate was 27.0 ($SD = 17.3$), 95% CI [20.2, 33.7]. In total, 17 patients (63%) scored above the cut-off of 16 symptoms. Endorsement rates were not normally distributed (Shapiro–Wilk $p < 0.05$). Mean score in patients with a poor proficiency was 33.4 ($SD = 19.4$). In patients with an intermediate proficiency this was 20.4 ($SD = 14.0$), while those with a good proficiency endorsed on average 26.5 symptoms ($SD = 17.6$). As indicated by a Kruskal–Wallis test, these group differences did not attain significance ($p = 0.24$). Mean symptom endorsement in patients with a positive incentive was 43.4 ($SD = 9.2$). Mean symptom endorsement in patients with a negative incentive was 11.7 ($SD = 4.4$), while those with "no" incentive attained a score of 15.2 ($SD = 10.2$). A Kruskal–Wallis test showed that this difference was significant, $p < 0.001$.

There were no missing forced-choice data and Cronbach's alpha for this test was 0.95. The mean error score was 12.4 ($SD = 11.8$), 95% CI [7.8, 17.0]. In total, 11 patients (41%) scored above the cut-off point of 9. The forced-choice data were not normally distributed (Shapiro–Wilk $p < 0.01$). The mean error score of patients with a poor proficiency was 17.7 ($SD = 14.7$). Those with an intermediate proficiency had on average 10.7 ($SD = 10.0$) errors and the error rate of those with a good proficiency was 7.8 ($SD = 7.5$). A Kruskal–Wallis test indicated that group differences in forced-choice errors were not significant: $p = 0.19$. Mean error score of patients with a positive incentive was 21.2 ($SD = 12.7$). Patients with a negative incentive to over-report symptoms had on average 6.8 errors ($SD = 3.3$). For patients with "no" incentive, the mean number of errors was 4.8 ($SD = 4.3$). A Kruskal–Wallis test indicated that these group differences were significant, $p < 0.01$.

Forced-choice errors and number of endorsed SIMS symptoms correlated at $r = 0.74$, $p < 0.01$. All patients who scored above the cut-off point on the SIMS, also did so for the forced-choice test.

The mean score on the dissociative items of the DES-II was 22.6 ($SD = 21.5$). Cronbach's alpha was 0.97; there were no missing data. DES-II scores correlated both with endorsed SIMS symptoms and forced-choice errors: $r = 0.79$, $p < 0.001$ and $r = 0.41$, $p < 0.05$, respectively. The DES-II item on difficulties to recognize oneself in a mirror was endorsed by nine patients (33%) for at least 10% of the time (range: 20%–100%). The reported mean frequency of this experience in all patients was 19.6% of the time ($SD = 32.6$). Eight patients (30%) endorsed the additional item on seeing images of past traumatic

experiences in the mirror. The reported mean frequency of this experience in all respondents was 16.9% of the time ($SD = 30.5$).

Table 2 gives rates of endorsed SIMS symptoms and forced-choice errors in the two symptom groups. As can be seen, the results replicate the pattern that is evident in Table 1. Thus, patients who had difficulties recognizing themselves in mirrors had higher SIMS scores and made more forced-choice errors than those who did not report this symptom (Mann–Whitney U: $p < 0.01$ and $p < 0.01$, respectively). Likewise, those who reported this symptom tended to fail more often on the SIMS (Fisher exact $p = 0.09$) and more often failed on the forced-choice task (Fisher exact $p = 0.01$). Patients who reported seeing aversive images in mirrors attained higher error scores on both the SIMS and the force-choice procedure than those who did not report this symptom (Mann–Whitney U: $p = 0.01$ and $p = 0.01$, respectively). Similarly, those who reported this symptom more often failed on the SIMS and the forced-choice task (Fisher exact p's $= 0.01$ and 0.01, respectively).

4. Discussion

The results of the current study can be summarized as follows. Firstly, we found that considerable proportions (27–37%) in samples of asylum applicants who were admitted to a psychiatric ward reported severe depersonalization symptoms. These are typically symptoms that would warrant asylum for medical reasons, but their prevalence is not out of line with rates found elsewhere in clinical settings (Foote, Smolin, Kaplan, Legatt, & Lipschitz, 2006; Friedl, Draijer, & de Jonge, 2000). Secondly, as a group, applicants had relatively high scores on our SVTs and many (41–87%) failed on these instruments when the conventional cut-points were employed. Thirdly, SVT scores were more strongly associated with incentives to malinger than with poor proficiency in the host language. That is, positive incentives to malinger were associated with higher endorsement of SIMS symptoms and more errors on our forced-choice task modeled after the MENT (Morel, 1998); poor proficiency in Dutch was only related to more errors on this forced-choice task — the SVT that was least dependent on interpretation.

There can be little doubt that asylum seekers form a highly vulnerable group, with high rates of psychopathology, notably trauma-related psychopathology (Reko et al., 2015). It is also clear that the psychopathology in this group is often evaluated in an unstructured way, i.e., without using standardized screening or diagnostic instruments, which raises diagnostic uncertainty (Maier, Schmidt, & Mueller, 2010).

Table 2
Number of endorsed SIMS items and forced-choice errors of patients in sample 3 ($N = 27$) who did (yes) or did not (no) report the target symptoms of depersonalization.

	Not recognizing oneself in the mirror		Seeing images from the past in the mirror	
	Yes ($n = 9$)	No ($n = 18$)	Yes ($n = 8$)	No ($n = 19$)
Mean (SD) SIMS	40.2 (15.2)	20.3 (14.6)	43.0 (12.5)	20.2 (14.5)
n (%) >16 SIMS	8 (89%)	9 (50%)	8 (100%)	9 (47%)
Mean (SD) forced-choice errors	21.1 (14.3)	8.1 (7.5)	24.1 (13.6)	7.5 (6.6)
n (%) >9 errors Forced-choice	7 (78%)	4 (22%)	7 (88%)	4 (21%)

Our results should not be taken to imply that there are many malingerers among asylum seekers. Rather they suggest that a tendency to over-report symptoms is prevalent in this group, which speaks to the need to evaluate symptomatology in this group in a standardized and more controlled way. A tendency to over-report symptoms may reflect many underlying factors, one of which is illness behavior so as to access healthcare services (McColl, McKenzie, & Bhui, 2008). That is, symptom over-reporting does not exclude the presence of real pathology. We do know, however, that over-reporting may obscure diagnostic evaluations and result in greater health care utilization (e.g., Horner, VanKirk, Dismuke, Turner, & Muzzy, 2014).

Several limitations of the current study should be noted. To begin with, the SVTs were introduced not for research purposes, but as an internal quality control during the implementation of ROM. Part of the staff who presented the tests were also involved in the treatment process, which means that blinding was not complete. However, data about the incentives and proficiency levels of the patients were gathered independently by other staff after the tests were completed.

Another limitation is that our samples were relatively small and therefore our results need replication, preferably with different SVTs. The SVTs that we used have been translated in other languages, but their validity has not been established for target groups with the highly diverse cultural backgrounds that are typical for asylum seekers in Western European countries. Because of this cultural diversity, we had to rely on an oral version of the SIMS that was translated in an ad hoc fashion by professional interpreters. Likewise, for giving the instructions of the forced-choice task to the patients, we relied on interpretation by professional interpreters. Thus, our SVTs were administered in a suboptimal way, and therefore mean scores and percentages of patients who failed these tests should be interpreted with caution. Relatedly, although translation of the items was done by professional interpreters, previous research has shown that interpreters might produce translation errors (Bot, 2005) and these may affect results.

The other side of the coin is that administering diagnostic tests with help of professional translators is routine practice in this clinic and in similar facilities for asylum seekers; our results are therefore a realistic reflection of diagnostic validity in clinical practice. Still, it would be a step forward if SVTs would be available that have pre-translated templates for various language groups, such that clinicians and researchers are less dependent on translators.

Although an ad hoc translation procedure was also used for the DES, this test – in contrast to the SVTs – has been tested in different cultural communities across several continents; its crosscultural validity is therefore considered to be good (Fan et al., 2011; Zoroglu, Sar, Tuzun, Tutkun, & Savas, 2002). The fact that prevalence of the "mirror sign" in the third sample of our study – where the DES was used – is comparable to its prevalence in the other samples does not support the notion that the cultural validity of the methods employed in these samples was inferior.

Unfortunately, in our samples, the various cultural and linguistic groups were represented by too few patients to run separate analyses for cultural subgroups. This is an important shortcoming of our study if only because traumatic experiences of Yugoslavian refugees, for example, may be expected to be tremendously different from those from East Asia or Africa.

A final limitation concerns the way in which the incentive potential was evaluated by social workers. The social workers inspected the files, thereby checking a restricted number of possible (positive and negative) incentives. Obviously, this is a crude approach, because incentives might be subtle (e.g., a desire to please the therapist) and need not to be documented in patient files. However, even with this crude approach we found deviant SVT performance to be related to the presence of positive incentives.

Our results, then, cast doubts on the validity of symptom reports about not being able to recognize oneself in mirrors and seeing aversive images in mirrors. Although the present analysis fails to confirm a large effect of language proficiency on SVT performance, we would overstate our case if we would argue that applicants who failed our SVTs are malingerers. For one thing, cultural issues may affect both symptom reports and performance on SVTs. What can be said with some confidence, though, is that – for whatever reason – the validity of these depersonalization symptoms cannot be taken at face value and that when asylum patients report them a thorough follow-up examination is in order. This is important because when therapists are under the impression that their asylum patients do have these symptoms, when in fact they do not, this may result in wrong treatment decisions (e.g., prescribing medication with potentially harmful side-effects).

5. Conclusions

Our study was exploratory in nature. Our attempt to differentiate between language proficiency levels and incentives to over-report might have been problematic because, for example, a tendency among the respondents to exaggerate cognitive impairment might have affected the evaluation of their proficiency in the host language. Still, our results suggest that when asylum applicants report during a free, unstructured psychiatric interview severe depersonalization symptoms, the validity of these symptoms cannot be taken at face value. Our study illustrates that this domain may benefit from the systematic administration of SVTs, although such instruments can by themselves never reveal whether or not a patient is intentionally over-reporting (Merten & Merckelbach, 2013). Several position papers have recommended the use of SVTs in neuropsychological assessment, emphasizing that validity testing is crucial for an accurate interpretation of clinical data (Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). We would argue that much the same is true for psychiatric evaluations in asylum seekers.

For future research, relating SVTs to culturally validated self-report instruments instead of open interviews may further clarify the relative impact of differences in culture, language, and secondary gain expectations in this target group.

References

Abély, P. (1927). Etat schizophrénique et tendances homosexuelles. *Annales Médico-Psychologiques, 2,* 251–257.

Artiola i Fortuny, L., & Mullaney, H. A. (1998). Assessing patients whose language you do not know: Can the absurd be ethical? *The Clinical Neuropsychologist, 12,* 113–126.

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease, 174*(12), 727–735.

Bot, H. (2005). *Dialogue interpreting in mental health.* New York: Rodopi.

Cirlugea, O. (2014). Assessing effort and malingering with the Asian client. In L. T. Benuto, N. S. Thaler, & B. D. Leany (Eds.), *Guide to psychological assessment with Asians* (pp. 165–180). New York: Springer.

Darwin, C. (1872). *The expression of the emotions in man and animals.* London: John Murray.

Drob, S. L., Meehan, K. B., & Waxman, S. E. (2009). Clinical and conceptual problems in the attribution of malingering in forensic evaluations. *The Journal of the American Academy of Psychiatry and the Law, 37,* 98–106.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–129.

Fan, Q., Yu, J., Ross, C. A., Keyes, B. B., Dai, Y., Zhang, T., ... Xiao, Z. (2011). Teaching Chinese psychiatrists to make reliable dissociative disorder diagnoses. *Transcultural Psychiatry, 48,* 473–483.

Foote, B., Smolin, Y., Kaplan, M., Legatt, M. E., & Lipschitz, D. (2006). Prevalence of dissociative disorders in psychiatric outpatients. *American Journal of Psychiatry, 163,* 623–629.

Friedl, M. C., Draijer, N., & de Jonge, P. (2000). Prevalence of dissociative disorders in psychiatric in-patients: The impact of study characteristics. *Acta Psychiatrica Scandinavica, 102,* 423–428.

Friedman, M. J., Resick, P. A., Bryant, R. A., Strain, J., Horowitz, M., & Spiegel, D. (2011). Classification of trauma and stressor-related disorders in DSM-5. *Depression and Anxiety, 28,* 737–749.

Gara, M. A., Vega, W. A., Arndt, S., Escamilla, M., Fleck, D. E., Lawson, W. B., et al. (2012). Influence of patient race and ethnicity on clinical assessment in patients with affective disorders. *Archives of General Psychiatry, 69,* 593–600.

Geraerts, E., Kozaric-Kovacic, D., Merckelbach, H., Peraica, T., Jelicic, M., & Candel, I. (2009). Detecting deception of war-related post traumatic stress disorder. *Journal of Forensic Psychiatry & Psychology, 20,* 278–285.

Goedhart, S., & Sno, H. N. (2014). Het 'signe du miroir'. Psychopathologische bespiegelingen aan de hand van een vergeten fenomeen [The 'signe du miroir'.

Psychopathological reflections on a forgotten phenomenon]. *Tijdschrift voor Psychiatrie, 56*, 328–335.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants 1 (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.

Horner, M. D., VanKirk, K. K., Dismuke, C. E., Turner, T. H., & Muzzy, W. (2014). Inadequate effort on neuropsychological evaluation is associated with increased healthcare utilization. *The Clinical Neuropsychologist, 28*, 703–713.

Kissane, M., Szymanski, L., Upthegrove, R., & Katona, C. (2014). Complex posttraumatic stress disorder in traumatized asylum seekers: A pilot study. *European Journal of Psychiatry, 28*, 137–144.

Lustig, S. L. (2008). Symptoms of trauma among political asylum applicants: Don't be fooled. *Hastings International and Comparative Law Review, 31*, 725–734.

Maier, T., Schmidt, M., & Mueller, J. (2010). Mental health and health care utilization in adult asylum seekers. *Swiss Medical Weekly, 140*, 13110.

McColl, H., McKenzie, K., & Bhui, K. (2008). Mental healthcare of asylum-seekers and refugees. *Advances in Psychiatric Treatment, 14*, 452–459.

Meffert, S. M., Musalo, K., McNiel, D. E., & Binder, R. L. (2010). The role of mental health professionals in political asylum processing. *Journal of the American Academy of Psychiatry and the Law Online, 38*, 479–489.

Meijer, D., & Noijons, J. (2008). *Gemeenschappelijk Europees referentiekader voor moderne vreemde talen: leren, onderwijzen, beoordelen [Common European Framework of reference for languages: Learning, teaching, assessment].* The Hague: Nederlandse Taalunie.

Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology, 18*, 145–152.

Merckelbach, H., Koeyvoets, N., Cima, M., & Nijman, H. (2001). De Nederlandse versie van de SIMS [The Dutch version of the SIMS]. *De Psycholoog, 10*, 586–591.

Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law, 6*, 122–137.

Montes, O., & Guyton, M. R. (2014). Performance of Hispanic inmates on the Spanish Miller Forensic Assessment of Symptoms Test (M-FAST). *Law and Human Behavior, 38*, 428–438.

Morel, K. R. (1998). Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *Journal of Personality Assessment, 70*, 299–314.

Morgan, C. (2007). Psychiatric evaluation of asylum seekers: Is it ethical practice or advocacy? *Psychiatry (Edgmont), 4*, 26–33.

Reko, A., Bech, P., Wohlert, C., Noerregaard, C., & Csillag, C. (2015). Usage of psychiatric emergency services by asylum seekers: Clinical implications based on a descriptive study in Denmark. *Nordic Journal of Psychiatry, 69*, 587–593.

Salazar, X. F., Lu, P. H., Wen, J., & Boone, K. B. (2007). The use of effort tests in ethnic minorities and in non-English-speaking and English as a second language populations. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 405–427). New York: Guilford.

Sollman, M. J., & Berry, D. T. R. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26*, 774–789.

Van IJzendoorn, M. H., & Schuengel, C. (1996). The measurement of dissociation in normal and clinical populations: Meta-analytic validation of the Dissociative Experiences Scale (DES). *Clinical Psychology Review, 16*, 365–382.

Van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist, 28*, 1336–1365.

Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology professional manual.* Odessa: Psychological Assessment Resources.

Yu-Fen, H., & Neng, T. (1981). Transcultural investigation of recent symptomatology of schizophrenia in China. *American Journal of Psychiatry, 138*, 1484–1486.

Zoroglu, S. S., Sar, V., Tuzun, U., Tutkun, H., & Savas, H. A. (2002). Reliability and validity of the Turkish version of the adolescent dissociative experiences scale. *Psychiatry and Clinical Neurosciences, 56*, 551–556.