ISSN 1577-7057

IJP&PT

INTERNATIONAL JOURNAL OF

PSYCHOLOGY & PSYCHOLOGICAL

THERAPY

https://www.ijpsy.com

IJP&PT is included in the following indexing and documentation Centers:
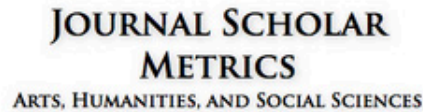
Applying the Questionnaire-based Implicit Association Test to Measure Automatic Negative Thinking

Francesco Dentale, Mara Stockner*, Michela Marchetti, Gianmarco Convertino, Giuliana Mazzoni

Sapienza Università di Roma, Italia

ABSTRACT

In a cognitive-behavioural theoretical framework, automatic negative thoughts are considered among the main determinants of depression and other disorders. Negative thinking is usually assessed through self-report scales, although several studies demonstrated their proneness to well-known confounds like introspective limits and social desirability. In the last decades several measures have been developed, within an implicit social cognition framework, that allow to moderate these confounds. Among them, the questionnaire-based Implicit Association Test (qIAT) is a latency-based paradigm well-suited to measure automatic propositional thinking. In this vein, two versions of the qIAT were designed to measure negative thinking, and successively tested in two different studies (with 118 and 71 participants, respectively). Internal consistency and concurrent validity of the new qIATs were assessed, along with their vulnerability to faking. Results showed adequate, even though not optimal, internal consistency for both qIATs. Across the two studies small/moderate positive correlations of the qIAT with two traditional self-report measures of depression were found, along with small/moderate negative correlations with satisfaction with life, self-esteem, and positive affects scales, supporting the concurrent validity of the new measures. Finally, both studies showed that the qIAT is considerably less vulnerable to faking compared to a traditional self-report scale of depression.
Key words: qIAT, depression, implicit social cognition, faking, negative thinking.

Novelty and Significance

What is already known about the topic?

- Self-report measures are prone to confounds like introspective limits and social desirability.
- New tools are needed in order to assess psychopathological symptoms, overcoming limitations of explicit measures.

What this paper adds?

- Propose a new propositional implicit measure to assess automatic negative thinking.
- Psychometric properties of the questionnaire-based Implicit Association Test are tested.
- The new tool is less prone to faking compared to a correspondent self-report questionnaire.

Automatic negative thoughts are key cognitive elements of depression (Beck, 1979; for a review see Colvin, Gardner, Labelle, & Santor, 2021) and in general are crucial in interpreting mental distress (Beck, 2008; Fenn & Byrne, 2013; Knapp & Beck, 2008). They comprise negative views about self, world and future (Beck, 1979). According to Beck (2002), automatic negative thoughts are particularly available in depressed individuals' memory, due to their biased cognitive schemas learned as the consequence of adverse experiences (Beck, 2008). The schemas can activate automatic information processing which is described as rapid and sparing attentional resources. On the other hand, the cognitive control system, responsible for problem solving and reappraisal is impaired in depression (Beck, 2008), leading to involuntary and automatically activated negative thoughts (Jurchis & Opre, 2018). This theoretical framework inspired several studies in

---

* Correspondence: Mara Stockner, Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza Università di Roma, Via degli Apuli 1, Roma, Italia. Email: mara.stockner@uniroma1.it

the literature (e.g, Mahali, Beshai, Feeney, & Misra, 2020), and it is considered the base of cognitive behavioural therapy (CBT) of depression (Gautam, Trpathi, Deshmukh, & Gaur, 2020). In this vein, one of the main practices of CBT for depression is to support patients in noticing and questioning their automatic negative thoughts, along with related dysfunctional beliefs (Cristea, Huibers, David, Hollon, Andersson, & Cuijpers, 2015; Cuijpers, Berking, Anderssson, Quigley, Kleiboe, & Dobson, 2013).

        The tendency to automatically activate negative thinking is considered responsible for the symptoms of depression and is assessed as a specific component in traditional self-report scales of depression, like the Beck Depression Inventory (BDI-II, Beck, Steer, & Brown 1996) or the Center for Epidemiologic Studies Depression Scale (CES-D, Radloff, 1977). Other more specific self-report scales have been developed to measure automatic negative thoughts (Harrell & Ryon, 1983; Hollon & Kendall, 1980). However, the validity of self-report scales to measure automatic negative thoughts is questionable for at least two main reasons. First, in accordance with the CBT theoretical framework, automatic negative thoughts can be made fully aware (and verbalized) only by activating specific introspective processes. The effectiveness of these introspective processes depends on individual capabilities and on situational conditions (e.g., the lack of distressing contextual factors). In this vein, any individual or situational factors that can impair mental introspection (e.g., Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005) may also biased responses on the self-report scales designed to measure automatic negative thinking. Second, self-report measures of psychopathological symptoms have been demonstrated to be vulnerable to faking, especially in sensitive assessment contexts (e.g., Rogers, 2008). Interestingly, a possible alternative way to investigate automatic negative thinking, attenuating the limitations of self-report scales, is given by the theoretical assumptions (and measurement instruments) of implicit social cognition.

        In the last decades different models of implicit social cognition have been proposed (see Payne & Gawronsky, 2010), providing a theoretical framework for the distinction between implicit and explicit cognition. In line with dual system models (e.g., Strack & Deutch, 2004), it is generally assumed that implicit (or spontaneous) evaluations depend on the automatic activation of mnemonic cognitive associations that are learned through evaluative conditioning processes, whereas explicit (or reflective) judgements depend on more informed processes based on deliberative thinking. Dual models assume that implicit representations are assessed normatively by deliberative processes that can confirm or disconfirm their contents (e.g., Gawronski & Bodenhausen, 2006). Differently from implicit representations, the coding format of explicit representations is not associative but propositional. In accordance with this theoretical framework, Beevers (2005) proposed a dual process model of cognitive vulnerability to depression. Negative associations towards the self are assumed to be among the main determinants of cognitive vulnerability to depression, as they make automatic negative thoughts particularly available in memory. The pathogenic power of these negative self-associations can be moderated by the deliberative (explicit) self-judgments that are able to disconfirm the former ones, at least when a sufficient degree of cognitive resources is available.

        Many instruments were developed to measure mental associations within this theoretical framework (e.g., Kurdi & Banaji, 2021). Among them, the classical Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is surely the most used and tested. Interestingly for the present research, various studies (e.g., Meites, Deveney, Steele, Holmes, & Pizzagalli, 2008; Dentale, Grano, Muzi, Pompili, Erbuto, & Violani, 2016) applied the classical IAT to measure depression in both healthy and

clinical samples, providing evidence for its reliability and validity. More specifically, the depression IAT is designed to investigate the automatic negative self-schema, measuring the degree of associations between "self vs. others" categories and "depression vs. wellness" attributes (Dentale *et alii*, 2016). Notwithstanding the interest of these studies, it is worth noting that the classical IAT showed an important limitation that undermines its validity in measuring automatic negative thoughts. As shown for the self-esteem IAT (Remue, Hughes, De Houwer, & De Raedt, 2014; Dentale, Vecchione, Ghezzi, Spagnolo, Szemenyei, & Barbaranelli, 2020), the exclusive use of single words as stimuli makes it possible to measure only the mere associations between self vs. other categories and depression vs. wellness attributes, making the results not interpretable. For instance, a strong association between "self" and "happiness" categories may be interpreted as "I am happy", or as "I want to be happy", or "I must be happy". In a similar way, a strong association between "self" and "sadness" categories may indicate that "I am sad", or that "I want to be sad", or again that "I must be sad" (to make the last two possibilities more realistic we can imagine for instance a faking-bad assessment scenario). It is worth noting that an associative measure like the depression IAT cannot disambiguate these possible interpretations of results. Moreover, since automatic negative thoughts (as conceived in the CBT perspective) are expressed in a propositional format like explicit beliefs and judgements, an associative implicit measure (like the depression IAT) can be only a proxy for their accurate assessment.

In contrast with the dual model framework, various authors (e.g., De Houwer, 2014) assumed that implicit (or spontaneous) evaluations depend on the automatic activation of propositional representations, and not on the activation of mental associations. Differently from mere associations (self-wellness vs. self-depression), propositional representations include information about the type of relationship between concepts in memory (I feel fine ≠ I want to feel fine ≠ I must feel fine) that are learned as a product of the interaction between the individual and the verbal community (Hayes, Barnes-Holmes, & Roche, 2001) At the same time, like mental associations, propositional representations can be formed and retrieved in an automatic manner from memory (De Houwer, 2014, 2018).

This different conceptualization of implicit cognition led to the development of new implicit measures, such as the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010) and the Relational Responding Task (RRT; De Houwer, Heider, Spruyt, Roets, & Hughes, 2015). Importantly for the present research, among variants of the IAT, De Houwer (2014) identified some paradigms that incorporate also relational information. In this category can be included both the autobiographical IAT (aIAT; Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008) and the questionnaire-based IAT (qIAT; Yovel & Friedman, 2013), as they use complete statements as stimuli, and no single words or images. In particular, the qIAT is designed to measure implicitly psychological constructs using as stimuli specific statements taken or adapted from the items of self-report measures. For example, Yovel and Friedman (2013) applied the qIAT to measure personality traits, such as introversion/extraversion. In the qIAT experimental paradigm, participants were instructed to complete various blocks of categorization tasks using two attributes (False vs. True) and two target categories (e.g., I am extraverted vs. I am introverted). Attribute categories (False vs. True) are stimuli/sentences which are either necessarily true, usually related to the testing situation (e.g., "I am sitting in front of a computer") or necessarily false (e.g., "I at the beach"). Target categories are stimuli/sentences related to the assessed topic, which indicate two possible opposite personality characteristics of the participants

(e.g., "I like to participate in parties" vs. "I like to be alone"). Participants are invited to categorize the stimuli shown on a computer screen, pressing two different keys on the keyboard, as quickly and accurately as possible. Both latencies and accuracy were recorded. By comparing mean reaction times of different blocks of trials, it is possible to know which individual attribute (e.g., extraversion vs. introversion) has a stronger association with True and which is more associated with False. Only few studies in literature explore the psychometric properties of the qIAT (Currie, Katz, & Yovel, 2017; Friedman, Katz, Cohen, & Yovel, 2021; Yovel, Aviram, Kahana, & Katz, 2021), but all of them showed adequate levels of internal consistency as well as evidence for its convergent, discriminant and predictive validity.

In line with propositional models, the qIAT seems to be a better alternative to measure automatic negative thoughts compared with the classical IAT, as the latter cannot include statements as stimuli, with their crucial intrinsic relational information. Instead, the propositional nature of the qIAT allows us to use (as items) statements taken from relevant self-report scales, by selecting the best items that measure negative ideation. At the same time, like for other variants of the Implicit Association Test (e.g., the classical IAT and the aIAT), the implicit nature of the qIAT should lead to greater resistance to faking effects, when compared to self-report measures, even if one cannot expect the qIAT to be completely immune to them (e.g., Fiedler & Bluemke, 2005; Verschuere, Prati, & Houwer, 2009).

The present research is aimed at applying the qIAT to measure automatic negative thinking, with stimuli inspired by the content areas of Beck's cognitive theory of depression. In particular, a first aim of the study was to assess the psychometric characteristics and the concurrent validity of a first version of the qIAT. Moreover, since the measurement of automatic negative thinking may be relevant also in sensitive assessment contexts (e.g., disability allowance application) a second aim of the study is to test the fakeability of the new qIAT, compared to a traditional self-report measure like the BDI-II.

In the second study, a new version of the qIAT was tested with a rewording of the stimuli-statements to prevent the use of undesirable recoding strategies during the categorization trials. As in the first study, study 2 was aimed to test the psychometric properties, the concurrent validity, and the vulnerability to faking of the new implicit measure.

Study 1

The first study is aimed at assessing the psychometric properties, concurrent validity, and vulnerability to faking effects of a questionnaire-based IAT (Yovel & Friedman, 2013) designed to measure automatic negative thinking with stimuli inspired by Beck's cognitive theory of depression, focusing on negative views of the self, the world and the future (for a complete list of the stimuli contact the corresponding author).

For the psychometric properties of the negative thinking qIAT, reliability was estimated in terms of internal consistency, while concurrent validity was analysed estimating the correlations with BDI-II and CES-D scales. To investigate the vulnerability to faking effects of the new measure, an experimental study was designed in which participants were randomly assigned to two groups: in the first they were invited to perform the qIAT and compile the BDI-II as honestly as possible, while in the second they were instructed to fake both the qIAT and the BDI-II in order to appear sufficiently depressed to obtain a disability allowance.

Method

Participants

One-hundred and eighteen undergraduate students, enrolled at Sapienza University of Rome (78 females; M_{age}= 24.36, SD= 3.78), participated in the present study. Twenty-four respondents had a bachelor's degree, 65 had a master's degree, while the remaining 29 were not graduates. Moreover, 108 participants were single, 9 were cohabiting, and one was married. They were recruited via social media as volunteers and did not receive rewards for participating. 59 participants (42 females; M_{age}= 24.58, SD= 4.03) took part in the no-faking condition whereas 59 other participants (36 females; M_{age}= 24.14, SD= 3.53) took part in the faking-condition.

Design and Procedure

The present study consisted of an experimental design with between-subject (no-faking vs. faking condition) and within-subject (implicit vs. explicit depression) measures. The experiment was carried out for all participants by simultaneously using Google Meet (to provide the instructions) and the software Inquisit (for the experimental procedure). All participants were randomly assigned to the no-faking or the faking condition.

At the beginning of the experimental session participants read and signed the informed consent form. In the no-faking condition, participants were invited to complete the BDI-II and the negative thinking qIAT as honestly as possible, while in the second condition, they were instructed to fake both these measures to appear depressed. In the faking condition, participants were invited to imagine a particular scenario: simulating a depression to obtain a disability allowance for which they have recently applied, with their requests being under evaluation. Since most participants might not necessarily know which symptoms indicate depression, a list of typical symptoms of depression was provided in the faking condition at (See Appendix A for the specific instructions presented to the subjects). After BDI-II and qIAT administration, a further measure of depression (CES-D; Radloff, 1977) was administered in an honest condition to all 118 participants. A socio-demographic questionnaire was finally administered. The whole procedure lasted around one hour.

Instruments and Measures

Socio-demographic Questionnaire. This questionnaire assessed the participants' age, gender, education level, relationship status and their previous/current experience with depression.

Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown 1996; Ghisi, Flebus, Montano, Sanavio, & Sica, 2006). The BDI-II is a scale that includes 21 items (each with four response options) designed to assess presence and severity of depressive symptoms in the last two weeks. It is one of the most used depression scales for clinical research, with adequate internal consistency and validity (Cronbach's α= .80; see Ghisi *et alii*, 2006).

Center for Epidemiologic Studies Depression Scale (CES-D; Fava, 1983; Radloff, 1977). The CES-D is a self-report scale that includes 20 items aimed to assess different facets of depression. Subjects were instructed to report the frequency of these different symptoms in the last week, using a 4-point Likert scale. Adequate levels of internal consistency (>.80) were found for this scale in different studies for both the original (Radloff, 1977) and the Italian version (Fava, 1983).

Negative Thinking Questionnaire-based Implicit Association Test (qIAT; Yovel & Friedman, 2013). The qIAT was specifically designed for assessing negatively oriented thinking in accordance with Beck's cognitive theory of depression. It included 7 blocks of

trials. In the first training block (20 trials), 10 statements linked to negative thinking or positive thinking were randomly presented in the centre of the screen. Half of them is focused on negative thinking (e.g., "In the last two weeks I felt like a failure"), while the remaining are focused on positive thinking (e.g., "In the last two weeks I felt fulfiled"). Each participant was instructed to categorize the stimuli as fast and accurately as possible into the 'negative thinking' category (E key on the keyboard) or into the 'positive thinking' category (I key on the keyboard). In the second training block (20 trials), the respondents were invited to categorize 10 statements that were necessarily "true" (5 items, e.g., "I am in front of a computer") or necessarily "false" (5 items, e.g., "I am running on a lawn"), respectively into true ("E" key) or false ("I" key) categories. Both the third block and fourth block consisted of 40 randomly presented trials, including both negative/positive and true/false statements. Respondents were invited to categorize the target statements into 'negative' and 'positive' thinking categories, and the attribute statements into 'true' or 'false' categories. In the fifth block of 20 trials, the position on the screen of target categories was inverted. The negative thinking category ("I" key) was placed on the right of the monitor, whereas the positive thinking category was on the left ("E" key). Finally, the sixth and seventh blocks consisted of 40 randomly presented trials, including both negative/positive and true/false statements, as in the fourth block, this time inverting on the screen the side of the categorization (as in the fifth block). The order of blocks 3 and 4, as well as that of blocks 6 and 7 was randomized across participants. All statements were presented at the centre of the monitor until the participants provided a response. In case of wrong response, a red cross appeared on the screen below the stimulus and remained on the screen until the right response was given. Inter-trial intervals were fixed at 250 ms. In accordance with Greenwald, Nosek, and Banaji's (2003) indications, the D2 scoring algorithm was used to calculate the qIAT scores. Response times above 10000 ms were deleted, and participants with a percentage of latencies faster than 300 ms were removed. Higher scores on the qIAT reflected higher negative thinking.

Data Analysis

To assess the reliability of the implicit and explicit measures the McDonald's Omega procedure was applied. To test the difference from zero of the qIAT mean scores, the *t*-student procedure was applied. To test the internal validity of the qIAT and its relationships with explicit measures, the Pearson correlation procedure was used. To test the effect of the faking/no-faking instructions on the implicit and explicit measures, the MANCOVA procedure was used. All the variables follow substantially the normal distribution with skewness and kurtosis values into the valid range (George & Mallery, 2019). All the statistical tests were conducted using a critical alpha level of .05.

RESULTS

Means and *SD*s of negative thinking qIAT, BDI-II and CES-D scales are reported in Table 1, both for the total sample and for each experimental group (faking vs. no faking). No substantial violations of skewness and kurtosis parameters' criteria were found, indicating that all measures' distributions were substantially normal. Moreover, eighty-seven (73.7%) of these participants did not report any episodes of depression in their life, and 101 did not report depression currently (85.6%). Six participants did not respond to depression-related questions. No significant difference emerged between the no-faking and faking groups in age and gender.

All the means of the qIAT scores (i.e., both the general mean score and also groups' mean scores) were negative and significantly different from zero [$t(117)= -13.78$, $p <.001$ for the total sample; $t(58)= -11.74$, $p <.001$ for the no-faking group; $t(58)=$

-8.16, p <.001 for the faking group], indicating a higher accuracy and lower latencies when negative thinking sentences were associated with false rather than true sentences, and positive thinking sentences were associated with true rather than false sentences.

As expected, CES-D mean scores were not significantly different between the faking and no-faking groups (as in this case all participants were invited to respond honestly), whereas for the BDI-II higher levels of depression emerged in the faking rather than in the no-faking condition, confirming that participants followed appropriately the instructions to fake depression in the experimental session. The mean score of the qIAT in the faking condition is only slightly higher than that of the no-faking condition (see the section 2.3.4 on vulnerability to faking).

Internal consistencies of all measures, estimated with McDonald's Omega procedure, are reported in Table 1, indicating an adequate (i.e., sufficient) level of reliability for the qIAT, and optimal levels for both the BDI-II and the CES-D.

Table 1. Descriptive statistics and reliability.

	Condition	General M	Groups' M	General SD	Groups' SD	McDonald's ω
qIAT	No Faking	-.35	-.40	.28	.26	.62
	Faking		-.31		.29	.58
BDI-II	No Faking	1.33	.65	.90	.50	.93
	Faking		2.01		.67	.95
CES-D	No Faking	1.86	1.88	.52	.54	.90
	Faking		1.84		.50	

Notes: BDI-II= Beck Depression Inventory II; CES-D= Center for Epidemiologic Studies Depression Scale; qIAT= Questionnaire-Based Implicit Association Test.

To evaluate the internal validity of the negative thinking qIAT, mean latency and error percentage were estimated (only for no-faking participants, to avoid possible distortions due to faking instructions), along with the correlations of the new measure with blocks' order, mean latencies and error percentage. The average latency (M= 1452.29 ms; SD=333.00) and error percentage (M= 5.48; SD= 4.75) of the negative thinking qIAT are in line with the literature (Yovel & Friedman, 2013; Yovel *et alii*, 2021). No significant correlations emerged between the qIAT and blocks' order, mean latencies, percentage of errors, supporting the internal validity of the new measure.

To investigate the concurrent validity of the new implicit measure, the correlations of the negative thinking qIAT with BDI-II and CES-D were calculated, including only no-faking participants (n= 59), to avoid distortions due to the experimental instructions. Significant correlations of moderate size emerged between the qIAT and both the BDI-II (r= .35, p <.01) as well as the CES-D (r= .28, p <.05), providing first evidence for the concurrent validity of the new tool.

As a preliminary step, to test possible mean differences on depression between faking and no-faking groups, an ANCOVA analysis was conducted, with condition (faking vs. no faking) as the independent variable, depression scores (as measured with CES-D scale in an honest condition for all participants) as the dependent variable, and age and sex as covariates. No significant effects emerged for age and sex covariates and, in addition, no significant differences in depression baseline levels were found between faking and no-faking groups (see Table 2).

Subsequently, to compare the vulnerability to faking of the new qIAT and the BDI-II depression scale, a MANCOVA analysis was conducted, including condition

Table 2. Intergroup mean differences in the CES-D.

	Condition	*M*	*SD*	*F* (1, 114)	*p*	*η²*
CES-D	No Faking	1.88	.54	.37	.55	.003
	Faking	1.84	.50			

Note: CES-D= Center for Epidemiologic Studies Depression Scale.

| | Faking | 1.86 | .44 | | | |

Notes: BDI-II= Beck Depression Inventory II; qIAT= Questionnaire-Based Implicit Association Test.

(faking vs. no-faking) as the independent variable, both qIAT's and BDI-II's scores as dependent variables, and CES-D depression scores, and age and sex as covariates. The results showed a significant multivariate effect for the CES-D [$F(2, 112)= 5.62$, $p <.01$, *partial eta squared*= .09] and for age [$F(2, 112)= 3.37$, $p <.05$, *partial eta squared*= .06] covariates, with no significant effect for sex. As expected, a significant and large multivariate effect of condition was found [$F(2, 112)= 84.25$, $p < 001$, *eta squared*= .60].

Regarding the univariate analyses, a significant and large effect of the experimental condition (faking vs. no-faking) was found for the BDI-II total score, but not for the qIAT (see Table 3), suggesting that while the BDI-II is extremely prone to faking, the new qIAT is less vulnerable. Moreover, among the covariate variables, both CES-D and age showed significant effects respectively on the BDI-II scores [$F(1, 113)= 11.09$, $p= .001$, *partial eta squared*= .09], and on the qIAT [$F(1, 113)= 4.47$, $p <.05$, *partial eta squared*= .04]. No other univariate significant effects were found.

Table 3. Faking effect on the novel qIAT and the total BDI-II.

	Condition	*M*	*SD*	*F* (1, 112)	*p*	*η²*
qIAT	No Faking	-.40	.26	2.72	.10	.02
	Faking	-.31	.29			
BDI-II	No Faking	.65	.50	166.37	<.001	.60
	Faking	2.01	.67			

Notes: BDI-II= Beck Depression Inventory II; qIAT= Questionnaire-Based Implicit Association Test.

DISCUSSION

The negative thinking qIAT showed an adequate (although not optimal) level of internal consistency. Mean latencies and error percentages of the qIAT were in line with the results of the literature, and no significant correlations emerged with blocks' order, mean latencies and error percentages, supporting the internal validity of the new tool. Moreover, the results provided first evidence for the concurrent validity of the new qIAT and showed that it is strongly less vulnerable to faking effects compared to the BDI-II. One possible limitation of the qIAT is the use of a standardized wording for the target statements that may have facilitated the use of recoding strategies during the categorization, allowing participants to perform the task without a complete reading of the sentences.

STUDY 2

As mentioned above, the target statements used as stimuli for the qIAT in Study 1 were all affirmative and with a similar syntactical structure. On the one hand, the standardized stimuli may have facilitated the categorization task but on the other hand,

it may have induced participants to use cognitive strategies in their performance. In particular, the location of the different elements of the sentences (nouns, pronouns, verbs, adverbs, adjectives, etc.) was substantially the same for all trials. This may have provided the opportunity to categorize the stimuli without a complete reading of the statements, undermining the validity of participants' performance.

To overcome this issue and improve the validity of the qIAT, a new version of the measure was proposed in Study 2, with a more varied wording of the stimuli/sentences, and the inclusion of both affirmative and negative formulations for the statements. These variations were aimed at forcing participants to read the entire statements when performing the task, instead of relying just on the presence of single sentence elements as decisive cues.

Study 2 was aimed at investigating the psychometric properties, the concurrent validity, and the vulnerability to faking effects of this new version of the negative thinking qIAT. As in Study 1, participants were randomly assigned into two groups: in the first they were invited to perform the qIAT and fill out the BDI-II as honestly as possible, while in the second group they were instructed to fake both the qIAT and the BDI-II to appear depressed (with the same instructions of Study 1).

Moreover, in this study the qIAT was administered a second time at the end of the study, in an honest condition for all participants. This 'honest' measure was used as a further covariate to improve the statistical tests, and to evaluate the concurrent validity of the new measure on the entire sample.

Method

Participants

Seventy-one students enrolled at Sapienza University of Rome (48 females, 2 participants that did not report their gender), with M_{age}= 24.71 (SD= 2.69) participated in the current study. Twenty-one respondents had a bachelor's degree, 32 had a master's degree, while the remaining 16 were not graduates (with two missing values). Moreover, 63 respondents were single, 4 were cohabiting, and 2 were married (and two missing values). Participants were again recruited via social media and did not receive any rewards for participating. The no-faking condition consisted of 38 participants (24 females; M_{age}= 24.18; SD= 2.15) whereas the faking condition consisted of 31 participants (24 females; M_{age}= 25.35; SD= 3.14).

Design and Procedure

The experimental research design of Study 2 was consistent with Study 1, however additional self-report measures (in order to test the concurrent validity of the qIAT) as well as a second administration of the qIAT, were carried out. The experiment was conducted in an online study (conducted with the use of both Google Meet and Inquisit), and participants were randomly assigned to a no-faking or a faking condition.

At the beginning of the experimental session, an informed consent was presented to each participant to be read and signed. As in Study 1, in the no-faking condition, participants were invited to complete the BDI-II and perform the negative thinking qIAT as honestly as possible, while respondents in the faking condition were instructed to fake both these measures to appear depressed. In the latter condition participants were invited to imagine the same sensitive scenario of study 1 (i.e., an official assessment

for receiving a disability allowance), receiving the same instructions and information. After administering the BDI-II and the qIAT, an additional scale of depression (i.e., the CES-D; Radloff, 1977) and other scales of related constructs (i.e., self-esteem, satisfaction with life, optimism, positive and negative affect, and impression management) were administered in an honest condition to all 71 participants. Subsequently, a second administration of the negative thinking qIAT was conducted in an honest condition for all participants, along with that of a socio-demographic questionnaire (see Study 1).

Instruments and Measures

Socio-demographic Questionnaire. The same questionnaire as in Study 1 was administered.
BDI-II. See the *Instruments and Measures* Section of Study 1 for a description of this scale.
CES-D. See the *Instruments and Measures* Section of Study 1 for a description of this scale.
Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965; Prezza, Trombaccia, & Armento, 1997). The RSES was applied to assess self-esteem. This is a self-report scale formed by 10 Likert items, ranging from 0 to 4. Various studies (e.g., Schmitt & Allik, 2005) demonstrated the adequate internal consistency (Cronbach's $\alpha= .81$) and test–retest reliability ($.85 < rtt < .88$) of the RSES. A global score of self-esteem was calculated averaging the score of all items.
Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985; Di Fabio & Ghizzani, 2007). The SWLS consists of 5 Likert items ranging from 1 (strongly disagree) to 7 (strongly agree). Both reliability (Cronbach's $\alpha= .87$; $rtt= .82$) and validity (Diener *et alii*, 1985) of the scale were demonstrated to be adequate. A global score of satisfaction with life was calculated averaging all items scores.
Life Orientation Test Revised (LOT-R; Scheier, Carvi & Bridges, 1994; Giannini, Schuldberg, Di Fabio & Gargaro, 2008). The LOT-R was applied to measure optimism. The LOT-R consists of 10 Likert items ranging from 1 (strongly disagree) to 5 (strongly agree). Four of the items were included as fillers and are not used to compute the global score. This scale revealed an adequate level of internal consistency (Cronbach's $\alpha= .82$; see Scheier *et alii*, 1994) and of test–retest correlation ($rtt= .79$; see Smith, Pope, Rhodewalt, & Poulton, 1989). A global score of optimism was calculated averaging all the items scores.
Positive and Negative Affect (SPANE; Diener *et alii*, 2009). The SPANE was applied to assess positive and negative affect. Respondents were instructed to indicate their feelings using 6 attributes of positive affect and six attributes of negative affect on a 1 to 5 points scale. Global scores for these scales were calculated averaging separately positive affect and negative affect items. Both scales showed an adequate internal consistency with alpha values $>.80$ (e.g., Diener *et alii*, 2009).
Impression Management (IM as part of the Balanced Inventory of Desirable Responding -BIDR; Paulhus, 1991). This scale consists of 20 seven-point Likert items aimed to measure the individuals' tendency to self-enhancement. Various studies support the reliability of the scale with an internal consistency ranging from .75 to .86, and with a test-retest correlation of .65 (Paulhus, 1991). In line with the indication of the author, the items were recoded as '1' if the score is 6 or 7, and as '0' if the score is lower or equal to 5. A general score of impression management was computed averaging the recoded items scores.

Data Analysis

To assess the reliability of the implicit and explicit measures, the McDonald's Omega procedure was applied. To test the difference from zero of the qIAT mean scores the *t*-student procedure was applied. To test the internal validity of the qIAT and its relationships with explicit measures, the Pearson correlation procedure was used. To test the effect of the faking/no-faking instructions on the implicit and explicit measures, the MANCOVA procedure was used. All the variables follow substantially the normal

distribution with skewness and kurtosis values into the valid range (George & Mallery, 2010). All the statistical tests were conducted using a critical alpha level of .05.

Results

Table 4 reports the descriptive statistics of all measures included in the study both for the total sample and for each experimental group (faking vs. no-faking). No substantial violations of skewness and kurtosis parameters' criteria were found, indicating that all measures' distributions were substantially normal. Regarding the participants' history of depression, fifty-seven (80.3%) of these participants did not report any episodes of depression in their life, and sixty-two did not report depression at that time (87.3%). Two participants did not respond to the questions on depression. No significant differences were found between no-faking and faking groups on these responses.

As in the first study, mean scores of the qIAT in both administrations were negative and significantly different from zero. This was true for the total sample and also for the experimental groups [Total sample, first administration: $t(70)= -11.27$, $p <.001$; Total sample, second administration: $t(70)= -7.97$, $p <.001$; No-faking group, first administration: $t(37)= -9.94$, $p <.001$; No-faking group, second administration: $t(37)= -5.29$, $p <.001$; Faking group, first administration: $t(32)= -6.25$, $p <.001$; Faking group, second administration: $t(32)= -5.99$, $p <.001$]. As in Study 1, a higher accuracy and lower latencies emerged when negative thinking sentences were associated with false sentences rather than with true sentences.

The internal consistencies of all measures, estimated with McDonald's omega procedure, are reported in Table 4. The internal consistency of the qIAT in the first and second administration indicated a moderate and an adequate level of reliability, respectively. The internal consistency of all other scales showed moderate to optimal levels, ranging from .74 to .91.

Table 4. Descriptive statistics and reliability.

	Group	General *M*	Groups' *M*	General *SD*	Groups' *SD*	McDonald's ω
qIAT	No Faking	-.32	-.36	.24	.23	.76
	Faking		-.27		.24	.59
qIAT2	No Faking	-.19	-.18	.21	.21	.63
	Faking		-.21		.20	
BDI-II	No Faking	1.18	.59	.77	.47	.92
	Faking		1.85		.43	.88
CES-D	No Faking	1.88	1.83	.56	.57	.91
	Faking		1.93		.56	
LOT	No Faking	3.21	.80	.87	.80	.86
	Faking		.95.		.95	
RSES	No Faking	2.90	2.91	.57	.60	.91
	Faking		2.89		.55	
SWLS	No Faking	4.41	4.35	1.31	1.37	.90
	Faking		4.47		1.25	
SPANE-POS	No Faking	3.35	3.37	.72	.78	.89
	Faking		3.33		.65	
SPANE-NEG	No Faking	2.52	2.46	.75	.74	.86
	Faking		2.58		.77	
IM	No Faking	.42	.41	.19	.21	.75
	Faking		.43		.17	

Notes: BDI-II= Beck Depression Inventory II; CES-D= Center for Epidemiologic Studies Depression Scale; IM= Impression Management Scale; LOT= Life Orientation Test; qIAT1= qIAT, first administration; qIAT2= qIAT, second administration; RSES= Rosenberg Self Esteem Scale; SPANE-POS= Scale of Positive and Negative Experience Positive Emotions; SPANE-NEG= Scale of Positive and Negative Experience Negative Emotions; SWLS= Satisfaction with Life Scale.

Mean latencies of the two qIAT were respectively 1502.97 ms (SD= 385.32) for the first qIAT (excluding participants in faking condition), and 1504.96 (SD= 418.93) for the second qIAT (total sample). The mean error percentage was 5.84 (SD= 4.39) for the first qIAT (excluding participants in faking condition), and 6.06 (SD= 5.03) for the second qIAT (total sample). Both these mean values are in line with results of the literature (Yovel & Friedman, 2013; Yovel *et alii*, 2021). Moreover, both qIATs were not significantly correlated with the order of blocks, mean latencies, and error percentage, supporting the internal validity of the test.

To evaluate the concurrent validity of the new implicit measure, the correlations between the negative thinking qIAT and the self-report scales included in the study were analysed (see Table 5). Since half of participants were invited to fake responses in both the first administration of the qIAT and in the BDI-II, these measures were excluded from the analysis as their correlations with the other scales are not interpretable in terms of concurrent validity.

Table 5. Correlations between the qIAT and the self-report measures.

	CES-D	LOT	RSES	SWLS	PO	SPANE-POS	SPANE-NEG	IM
qIAT2	.24*	-.19	-.24*	-.36**	-.30*	-.28*	.04	-.03

Notes: CES-D= Center for Epidemiologic Studies Depression Scale; IM= Impression Management Scale; LOT= Life Orientation Test; qIAT2= qIAT, second administration; RSES= Rosenberg Self Esteem Scale; SPANE-POS= Scale of Positive and Negative Experience Positive Emotions; SPANE-NEG= Scale of Positive and Negative Experience Negative Emotions; SWLS= Satisfaction with Life Scale; PO= Positive Orientation; *= Correlation is significant at the .05 level (2-tailed); **= Correlation is significant at the .01 level (2 tailed).

As in study 1, the qIAT (second administration) showed a significant and positive correlation with depression (CES-D). Moreover, it showed small/moderate negative correlations with self-esteem (RSES), satisfaction with life (SWLS), and positive affect (Positive subscale of the SPANE), as well as with an estimated index of Positive Orientation (PO) calculated as a common factor of LOT, RSES and SWLS scores (Caprara, Alessandri, Abela, & McWhinnie, 2010), supporting concurrent validity of the qIAT. Correlations with optimisms (LOT), and negative affect (the Negative subscale of the SPANE), were negative but non-significant. Finally, also the correlation with impression management (IM) was non-significant, suggesting that qIAT scores are not influenced by personality tendencies to preserve social self-image.

Moreover, as in the first study, possible mean differences between faking and no-faking groups on the baseline levels of depression were tested. To this aim, initially an ANCOVA analysis was conducted, including the condition (faking vs. no faking) as an independent variable, depression scores (measured with CES-D scale) as a dependent variable, and age and sex as covariates. No significant effects emerged for age and sex covariates. Moreover, as expected, no significant difference in CES-D mean scores was found between faking and no-faking groups given that participants are not instructed to fake in CES-D (see Table 6).

Subsequently, to compare the vulnerability to faking of the new implicit qIAT measure and the BDI-II depression scale, a MANCOVA analysis was conducted, including the condition (faking vs. no-faking) as an independent variable, and both

Table 6. Intergroup mean differences in the CES-D.

	Condition	M	SD	$F (1, 65)$	p	η^2
qIAT	No Faking	1.83	.57	.12	.74	.00
	Faking	1.92	.55			

Notes: CES-D: Center for Epidemiologic Studies Depression Scale (honest completion for both conditions).

qIAT1's (first administration) and BDI-II's scores as dependent variables. The qIAT2 (second administration) and CES-D's scores, along with age and sex, were also included as covariates. The results showed a significant multivariate effect for the CES-D [F(2, 62)= 14.92, p <.001, *partial eta squared*= .33] and for the qIAT2 [F(2, 62)= 3.09, p= .05, *partial eta squared*= .09] covariates, while the effects for sex and age were not significant. As expected, a significant and large multivariate effect of condition [F(2, 62)= 95.04, p <.001, *eta squared*= .75] was found.

At the univariate level analysis, a significant and large effect of the experimental condition (faking vs. no-faking) was found for the BDI-II scores. A significant effect of condition was found also for the qIAT1, but with an effect size 9 times lower with respect to the BDI-II (see the last column of Table 7). These results confirm that the qIAT is substantially less vulnerable to faking than the BDI-II. Moreover, among the covariate variables, CES-D showed a significant effect on the BDI-II total score [F(1, 63)= 23.95, p< .001, *partial eta squared*= .28], while the qIAT2 revealed a close-to-significant effect on the qIAT1 scores [F(1, 63)= 3.73, p= .06, *partial eta squared*= .06]. No other significant univariate effects were found.

Table 7. Faking effect on the negative thinking qIAT and the total BDI-II.

	Condition	*M*	*SD*	*F* (1, 63)	*p*	η^2
qIAT	No Faking	-.36	.23	5.15	.03	.08
	Faking	-.24	.23			
BDI-II	No Faking	.59	.47	163.20	<.001	.72
	Faking	1.86	.44			

Notes: BDI-II= Beck Depression Inventory II; qIAT= Questionnaire-Based Implicit Association Test.

DISCUSSION

Overall, the new version of the negative thinking qIAT showed adequate levels of internal consistency in both administrations. Mean scores of latencies, and of error percentage, of both the qIATs are in line with other results reported in literature. No significant correlations emerged with blocks' order, mean latencies, and error percentage, confirming the adequate internal validity of the new measure. Moreover, the new qIAT showed significant and small/moderate correlations with the CES-D as well as with RSES, SWLS, POS and SPANE positive affect, providing first evidence for the concurrent validity of the new tool. As expected, results showed also that the impression management scale was not correlated with the qIAT, indicating that it was not influenced by individual tendencies to preserve social self-image.

Finally, the results confirmed that the qIAT is strongly less prone to faking effects compared to the BDI-II scale.

GENERAL DISCUSSION

In the present study the qIAT was applied to measure automatic negative thinking, providing first evidence for its reliability, as well as for its internal and concurrent validity. Moreover, the vulnerability to faking of the new implicit measure was tested and compared with that of a traditional self-report measure of depression (i.e., the BDI-II).

The results showed an adequate level of internal consistency, and mean latencies and error percentages in line with previous studies (Yovel & Friedman, 2013; Yovel

et alii, 2021). Moreover, the negative thinking qIAT was not significantly correlated with the order of blocks, mean latencies, and error percentages, supporting the internal validity of the new measure. Significant positive correlations of small/moderate size were found between the negative thinking qIAT and self-report scales of depression (i.e., BDI-II and CES-D), providing initial support for the concurrent validity of the new tool. These results are partially in line with the correlations found between the classical IATs and self-report scales of depression (Creemers, Scholte, Engels, Pieters, & Wiers, 2013; Van Tuijl, Glashouwer, Elgersma, Bockting, Penninx, & de Jong, 2018), even if not all previous studies confirmed these findings (Price, Panny, Degutis, & Griffo, 2021). The negative thinking qIAT showed also significant negative correlations of small/moderate size with self-report measures of satisfaction with life, self-esteem, positive oriented thinking, and positive affect, with a further support for its concurrent validity. Interestingly, similar results were found in previous studies (Creemers *et alii*, 2013; Dentale *et alii*, 2016) that revealed significant correlations between classical IATs and self-report measures of self-esteem and satisfaction with life. However, it is worth noting that the inclusion of entire propositions as stimuli represent a better tool to measure automatic thoughts, compared to the classical IAT. The classical IAT does not include relational information between conceptual categories (e.g., Self) and attributes (e.g., Happy), but only mere associative patterns (e.g., Remue *et alii*, 2014). This represents a strong limitation of the classical IAT, as it cannot disambiguate different possible interpretations of a strong association between two words. Consider the pair of concepts self and happy. A strong association identified by the IAT has a very ambiguous meaning, as it can indicate the current state of matters (I am happy), a hoped-for state of matters (I want to be happy) or even some imperative to obtain an idealized state of matters (I must be happy). Differently from the classical IAT, by presenting statements that include relational information among the terms of the sentence (e.g. I am happy vs I want to be happy vs I must be happy), the qIAT allows a nonambiguous interpretation of the person's internal condition (e.g., De Houwer, 2014). Moreover, when compared to self-report measures, an implicit measure like the qIAT may be more suited to detect and measure genuine automatic thoughts. Indeed, self-report measures are intrinsically based on introspective processes that cannot guarantee to detect spontaneous form of thinking (e.g., Hofmann *et alii*, 2005).

Another interesting result is that the qIAT appears not to be influenced by personality tendencies to preserve social self-image and is also less vulnerable to faking effects compared to self-report measures, like the BDI-II. This is in line with results in the existing literature, showing that self-report measures are vulnerable to faking (e.g., Grieve & Mahar, 2010) while implicit propositional measures are substantially less vulnerable, even if not completely immune to faking (Verschuere *et alii*, 2009; Dentale *et alii*, 2020). Notably, implicit measures were found to be particularly prone to faking when participants had previous experience with the IAT, and when participants received specific instructions or trainings on how to manipulate responses (e.g., Hu & Rosenfeld, 2012). These results support the necessity to use measures like the negative thinking qIAT in sensitive assessment contexts (when faking can be used to achieve secondary gains, as remuneration or avoiding work), with the caveat of proposing it to groups that are not experts in the use of the tool, nor are trained in how to respond and fake. Additional studies on this issue, however, are necessary in order to clarify the contextual conditions of faking in implicit tasks.

Also, in both studies, the internal consistency of the qIAT is adequate, but not optimal (McDonald's Omega ranging from .58 to .76). Internal consistencies are lower than those found in other studies (e.g., Friedman *et alii*, 2021; Yovel & Friedman, 2013), in which the qIAT was used to assess personality factors (with split-half correlations ranging from .77 to .89). Notably, in the case of the present research, the not optimal reliability may have reduced the correlations with the criteria used to assess the concurrent validity. To minimize a possible relevant source of casual error, and improve the reliability of the new measure, it might be useful in the future to avoid the online administration used for the present studies and organize single experimental sessions in the laboratory. This would increase the experimental control on the active participation of the participants, reduce casual answering and avoid technological burden linked, for instance, to the download of experimental software. Regarding the generalizability of our results, one limitation of both our studies is the sample size, which limits the applicability of our results in applied contexts. Given our seminal evidence, we suggest further studies should replicate our findings in a larger sample, controlling at the same time also for limitations related to self-report measures (e.g., social desirability). As the no-faking condition represented an "honest" condition for the present research, these could have impacted our results, limiting their validity. A further limitation is given by the lack of a clinically depressed sample, and future research should also consider including a sample of depressed patients. The administration of the automatic negative thinking qIAT to clinical samples would allow investigating the capacity of the negative thinking qIAT to discriminate between individuals who suffer from clinically recognized depression and non-clinical, healthy participants. If the discriminating power of the qIAT will be confirmed, it would provide further support for the criterion validity of the new measure. Moreover, in the current studies, proneness to faking of the new implicit measure was tested only by inviting participants to manipulate responses to appear more depressed than they really were. No specific information or training were provided, possibly limiting the participants ability to fake depression in the implicit measure. Several previous studies (e.g., Hu & Rosenfeld, 2012) demonstrated that specific instructions and trainings can considerably increase the capacity of participants to manipulate their scores on implicit measures. It might then be important in future studies to assess the effect of different instructions on the ability to fake in the negative thinking qIAT. Moreover, it may be interesting to apply some more sophisticated strategies, devoted to identify faking responses, based on individuals' IAT response-patterns (e.g., Agosta, Ghirardi, Zogmaister, Castiello, & Sartori, 2011), or on mathematical procedures designed to decompose IAT scores (e.g., Röhner & Ewers, 2016).

In our two studies, negative ideation was measured using exclusively the qIAT, even if also other interesting implicit propositional measures are available, such as the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et alii, 2010), the Relational Responding Task (RRT; De Houwer *et alii*, 2015), and the Propositional Evaluation Paradigm (PEP; Müller & Rothermund, 2019), along with its mouse-tracking version (MT-PEP; Cummins & De Houwer, 2021). Interestingly, the latter one was demonstrated to be more sensitive to relational information if compared with the RRT and the aIAT (Cummins & De Houwer, 2022).

Finally, the target stimuli of the negative thinking qIAT were inspired by the conceptual areas theorized by Beck (i.e., negative views of the self, world, and future). To improve the content validity of the new measure, it may be also important to refer to other theorizations (e.g., Harrell & Ryon, 1983; Hollon & Kendall, 1980).

Even if future studies should address the potential limitations mentioned before, the present research provides the first available evidence supporting the reliability as well as the internal and concurrent validity of an implicit measure of negative thinking based on the qIAT. The results also showed a strongly lower vulnerability to faking for the qIAT with respect to a traditional self-report measure of depression, as the BDI-II. Overall, these results encourage to further explore in future studies the potentialities of the qIAT as a measure of automatic negative thinking.

REFERENCES

Agosta S, Ghirardi V, Zogmaister C, Castiello U, & Sartori G (2011). Detecting fakers of the autobiographical IAT. *Applied Cognitive Psychology, 25*, 299-306. Doi: 10.1002/acp.1691

Barnes-Holmes D, Barnes-Holmes Y, Stewart I, & Boles S (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*, 527-542. Doi: 10.1007/BF03395726

Beck AT (2002). Cognitive models of depression. *Clinical Advances in Cognitive Psychotherapy: Theory and application, 14*, 29-61.

Beck AT (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry, 165*, 969-977.

Beck AT (1979). *Cognitive therapy of depression*. New York: Guilford Press.

Beck AT, Steer RA, & Brown GK (1996). *Beck Depression Inventory, 2nd edition*. San Antonio, TX: The Psychological Corporation.

Beevers CG (2005). Cognitive vulnerability to depression: A dual process model. *Clinical Psychology Review, 25*, 975-1002.

Caprara GV, Steca P, Alessandri G, Abela JR, & McWhinnie CM (2010). Positive orientation: Explorations on what is common to life satisfaction, self-esteem, and optimism. *Epidemiology and Psychiatric Sciences, 19*, 63-71.

Colvin E, Gardner B, Labelle PR, & Santor D (2021). The automaticity of positive and negative thinking: a scoping review of mental habits. *Cognitive therapy and research*, 1-27.

Creemers DH, Scholte RH, Engels RC, Pieters S, & Wiers RW (2013). Acute stress increases implicit depression and decreases implicit self-esteem. *Journal of Experimental Psychopathology, 4*, 118-132. Doi: 10.5127/jep.025411

Cristea IA, Huibers MJ, David D, Hollon SD, Andersson G, & Cuijpers P (2015). The effects of cognitive behavior therapy for adult depression on dysfunctional thinking: A meta-analysis. *Clinical Psychology Review, 42*, 62-71.

Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, & Dobson K S (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry, 58*, 376-385.

Cummins J & De Houwer J (2021). The Shape of Belief: Developing a Mousetracking-Based Relational Implicit Measure. *Social Psychological and Personality Science, 12*, 1517-1526. Doi: 10.1177/1948550620978019

Cummins J & De Houwer J (2022). Are relational implicit measures sensitive to relational information? *Collabra: Psychology, 8*, 38621.

Currie CJ, Katz BA, & Yovel I (2017). Explicit and implicit shame aversion predict symptoms of avoidant and borderline personality disorders. *Journal of Research in Personality, 71*, 13-16. Doi: 10.1016/j.jrp.2017.08.006

De Houwer J (2014). A propositional model of implicit evaluation. *Social and personality psychology compass, 8*, 342-353. Doi: 10.1111/spc3.12111

De Houwer J (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin, 13*, 1-21.

De Houwer J, Heider N, Spruyt A, Roets A, & Hughes S (2015). The relational responding task: Toward a new implicit measure of beliefs. *Frontiers in Psychology, 6*, 319. Doi: 10.3389/fpsyg.2015.00319

Dentale F, Grano C, Muzi M, Pompili M, Erbuto D, & Violani C (2016). Measuring the automatic negative self-schema: New evidence for the construct and criterion validity of the Depression Implicit Association. *Self and Identity, 15*, 599-613. Doi: 10.1080/15298868.2016.1179218

Dentale F, Vecchione M, Ghezzi V, Spagnolo G, Szemenyei E, & Barbaranelli C (2020). Beyond an Associative Conception of Automatic Self-Evaluations: Applying the Relational Responding Task to Measure Self-Esteem.

Psychological Record, *70*, 227-242. Doi: 10.1007/s40732-020-00392-4

Di Fabio A & Ghizzani F (2007). La soddisfazione di vita in un campione di apprendisti maggiorenni: Alcuni correlati e predittori. *Giornale Italiano di Psicologia dell'Orientamento*.

Diener ED, Emmons RA, Larsen RJ, & Griffin S (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71-75. Doi: 10.1207/s15327752jpa4901_13

Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, & Biswas-Diener R (2009). Scale of positive and negative experience. *PsycTESTS*. Doi: 10.1037/t03125-000

Fava GA (1983). Assessing depressive symptoms across cultures: Italian validation of the CES-D self-rating scale. *Journal of Clinical Psychology, 39*, 249-251. Doi: 1097-4679(198303)39:2<249

Fenn K & Byrne M (2013). The key principles of cognitive behavioural therapy. *InnovAiT*, *6*, 579-585.

Fiedler K & Bluemke M (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology, 27*, 307-316.

Friedman A, Katz BA, Cohen IH, & Yovel I (2021). Expanding the scope of implicit personality assessment: an examination of the questionnaire-based implicit association test (qIAT). *Journal of Personality Assessment, 103*, 380-391. Doi: 10.1080/00223891.2020.1754230

Gautam M, Tripathi A, Deshmukh D, & Gaur M (2020). Cognitive Behavioral Therapy for Depression. *Indian Journal of Psychiatry, 62*, 223-229.

Gawronski B & Bodenhausen GV (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692.

George D & Mallery P (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. New York: Routledge.

Ghisi M, Flebus GB, Montano A, Sanavio E, & Sica C (2006). *Beck depression inventory-II*. Firenze: Giunti OS.

Giannini M, Schuldberg D, Di Fabio A, & Gargaro D (2008). Misurare l'ottimismo: Proprietà psicometriche della versione Italiana del Life Orientation Test-Revised (LOT-R) [Measuring optimism: Psychometric properties of the Italian version of the Life Orientation Test-Revised (LOT-R)]. *Counseling: Giornale Italiano di Ricerca e Applicazioni*, *1*, 73-84.

Greenwald AG, McGhee DE, & Schwartz JL (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*, 1464. Doi: 10.1037/0022-3514.74.6.1464

Greenwald AG, Nosek BA, & Banaji MR (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197. Doi: 10.1037/0022-3514.85.2.197

Grieve R & Mahar D (2010). The role of fluid and emotional intelligence in malingering. *Australian Journal of Psychology, 62*, 103-111. Doi: 10.1080/00049530903032836

Harrell TH & Ryon NB (1983). Cognitive-behavioral assessment of depression: clinical validation of the automatic thoughts questionnaire. *Journal of Consulting and Clinical Pychology, 51*, 721.

Hayes SC, Barnes-Holmes D, & Roche B (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Springer.

Hofmann W, Gawronski B, Gschwendner T, Le H, & Schmitt M (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measure. *Personality and Social Psychology Bulletin, 31*, 1369-1385. Doi: 10.1177/0146167205275613

Hollon SD & Kendall PC (1980). Cognitive self-statements in depression: Development of an automatic thoughts questionnaire. *Cognitive Therapy and Research*, *4*, 383-395.

Hu X & Rosenfeld JP (2012). Combining the P300-complex trial-based concealed information test and the reaction time-based autobiographical implicit association test in concealed memory detection. *Psychophysiology, 49*, 1090-1100. Doi: 10.1111/j.1469-8986.2012.01389.x

Jurchiș R & Opre A (2018). A dual process approach to current issues from cognitive behavior therapies: An integrative perspective. *International journal of Cognitive Therapy*, *11*, 374-409.

Knapp P & Beck AT (2008). Cognitive therapy: foundations, conceptual models, applications and research. *Brazilian Journal of Psychiatry, 30*, s54-s64.

Kurdi B & Banaji MR (2021). *Implicit Social Cognition: A brief (and gentle) introduction*. Doi: 10.31234/osf.io/a4pjy

Mahali S, Beshai S, Feeney JR, & Mishra S (2020). Associations of negative cognitions, emotional regulation, and depression symptoms across four continents: International support for the cognitive model of depression. *BMC Psychiatry, 20*, 1-12. Doi: 10.1186/s12888-019-2423-x

Meites TM, Deveney CM, Steele KT, Holmes AJ, & Pizzagalli DA (2008). Implicit depression and hopelessness in remitted depressed individuals. *Behaviour Research and Therapy, 46*, 1078-1084.

Müller F & Rothermund K (2019). The Propositional Evaluation Paradigm: Indirect Assessment of Personal Beliefs and Attitudes. *Frontiers in Psychology*, *7*, Article 2385. Doi: 10.3389/fpsyg.2019.02385.

Müller F & Rothermund (2019). The propositional evaluation paradigm: Indirect assessment of personal beliefs and attitudes. *Frontiers in Psychology, 10,* Article 2385. Doi: 10.3389/fpsyg.2019.02385

Paulhus DL (1991). Measurement and control of response bias. In JP Robinson, PR Shaver, & LS Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). New York: Academic Press.

Payne BK & Gawronski B (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going. In B Gawronski & BK Payne (Eds.) *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1-15). New York: Guilford Press.

Prezza M, Trombaccia FR, & Armento L (1997). La scala dell'autostima di Rosenberg: Traduzione e validazione Italiana. Firenze: Giunti OS.

Price RB, Panny B, Degutis M, & Griffo A (2021). Repeated measurement of implicit self-associations in clinical depression: Psychometric, neural, and computational properties. *Journal of Abnormal Psychology, 130*, 152-165. Doi: 10.1037/abn0000651

Radloff LS (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401. Doi: 10.1177/014662167700100306

Remue J, Hughes S, De Houwer J, & De Raedt R (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *PloS one, 9*, e108837.

Rogers R (2008). Detection strategies for malingering and defensiveness. *Clinical Assessment of Malingering and Deception, 3*, 14-35.

Röhner J & Ewers T (2016). Trying to separate the wheat from the chaff: Construct- and faking-related variance on the Implicit Association Test (IAT). *Behavior Research Mmethods, 48*, 243-258. Doi: 10.3758/s13428-015-0568-1

Rosenberg M (1965). *Society and the adolescent self-image*. Princeton NJ: Princeton University Press.

Sartori G, Agosta S, Zogmaister C, Ferrara SD, & Castiello (2008). How to accurately detect autobiographical events. *Psychological Science, 19*, 772-780. Doi: 10.1111/j.1467-9280.2008.02156.x

Scheier MF, Carver CS, & Bridges MW (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*, 1063. Doi: 10.1037/0022-3514.67.6.1063

Schmitt DP & Allik J (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89*, 623. Doi: 10.1037/0022-3514.89.4.623

Smith TW, Pope MK, Rhodewalt F, & Poulton JL (1989). Optimism, neuroticism, coping, and symptom reports: an alternative interpretation of the Life Orientation Test. *Journal of Personality and Social Psychology, 56*, 640. Doi: 10.1037/0022-3514.56.4.640

Strack F & Deutsch R (2004). Reflective and impulsive determinants of social behavior. *Personality and social Psychology Review, 8*, 220-247.

van Tuijl LA, Glashouwer KA, Elgersma HJ, Bockting CL, Penninx BW, & de Jong PJ (2018). Depression recurrence after recovery: Prognostic value of implicit and explicit self-depressed associations. *Behaviour Research and Therapy, 107*, 76-82. Doi: 10.1016/j.brat.2018.06.001

Verschuere B, Prati V, & Houwer JD (2009). Cheating the lie detector: Faking in the autobiographical Implicit Association Test. *Psychological Science, 20*, 410- 413. Doi: 10.1111/j.1467-9280.2009.02308.x

Yovel I & Friedman A (2013). Bridging the gap between explicit and implicit measurement of personality: The questionnaire-based implicit association test. *Personality and Individual Differences, 54*, 76-80. Doi: 10.1016/j.paid.2012.08.015.

Yovel I, Aviram G, Kahana N, & Katz BA (2021). Testing a new indirect measure of general self-worth: The Self-esteem Questionnaire-based Implicit Association Test. *British Journal of Social Psychology*. Doi: 10.1111/bjso.12472

Appendix

Study 1: Instructions for the faking group

Imagine being examined by an insurance policy board to receive compensation for psychological harm. You have to make them believe that the damage caused you severe depression. Symptoms of depression are outlined below. Please read carefully.

1. Depressed mood for most of the day, most days.
2. Significantly decreased interest or pleasure in all, or nearly all, activities for most of the day, most days.
3. Significant weight loss (when not on a diet) or weight gain or decreased or increased appetite almost every day.
4. Insomnia or hypersomnia most days.
5. Agitation or psychomotor retardation most days.
6. Fatigue or loss of energy almost every day.
7. Feelings of worthlessness or excessive or inappropriate guilt almost every day.
8. Decreased ability to think or concentrate, or indecisiveness, almost every day.
9. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for suicide.

After having read these instructions, you will need to perform the same categorization task you just completed to assess depression. Your job is to answer by trying to simulate depression. In doing this, however, you must try to be credible, avoiding being exposed in your manipulation attempts and bearing in mind that if response times and / or errors increase too much, the task will be annulled.

Study 1: Instructions for the no faking group

After having read these instructions, your task is to carry out two measures, both of them aim to evaluate depression. In particular, you will first have to repeat the categorization task you just performed, and then you will also have to fill out a questionnaire.

In both cases, we ask you to carry out the tasks by answering as honestly as possible, paying the utmost attention to the instructions presented.

Study 2: Instructions for the faking group (BDI-II)

As a first task we ask you to carry out a test aimed to assess depression. When answering, however, we ask you NOT to be honest, but to imagine yourself being examined by an insurance policy board to receive compensation for psychological damage. Your goal is to make them believe that the damage caused you severe depression. Symptoms of depression are outlined below. Please read carefully.

1. Depressed mood for most of the day, most days.
2. Significantly decreased interest or pleasure in all, or nearly all, activities for most of the day, most days.
3. Significant weight loss (when not on a diet) or weight gain or decreased or increased appetite almost every day.
4. Insomnia or hypersomnia most days.
5. Agitation or psychomotor retardation most days.
6. Fatigue or loss of energy almost every day.
7. Feelings of worthlessness or excessive or inappropriate guilt almost every day.
8. Decreased ability to think or concentrate, or indecisiveness, almost every day.
9. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for suicide.

After having read these instructions, you will have to fill out a questionnaire aimed to assess depression, your task is to answer by simulating depression, trying to be credible and avoiding being unmasked.

Study 2: Instructions for the no-faking group

After having read these instructions, your task is to carry out three tests that will be presented to you IN THE MOST HONEST AND ACCURATE WAY POSSIBLE. In particular, you will be asked to fill out a questionnaire, and also to perform two timed categorization tasks. All three of these tasks are aimed to assess depression. The tasks will be explained as you continue with the administration. For all three tasks, please complete the proposed tests by answering as honestly as possible, paying the utmost attention to the presented instructions.

Study 2: Instructions for the faking group (qIAT)

The next task will not be a questionnaire, but a timed categorization task always aimed to assess depression. Similarly to the previous task, we ask you to imagine being examined by an insurance policy board to receive compensation for psychological harm. You have to make them believe that the damage caused you severe depression. Symptoms of depression are outlined below. Please read carefully.
Symptoms of depression are outlined below. Please read carefully.

1. Depressed mood for most of the day, most days.
2. Significantly decreased interest or pleasure in all, or nearly all, activities for most of the day, most days.
3. Significant weight loss (when not on a diet) or weight gain or decreased or increased appetite almost every day.
4. Insomnia or hypersomnia most days.
5. Agitation or psychomotor retardation most days.
6. Fatigue or loss of energy almost every day.
7. Feelings of worthlessness or excessive or inappropriate guilt almost every day.
8. Decreased ability to think or concentrate, or indecisiveness, almost every day.
9. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for suicide.

After having read these instructions, you will need to carry out the timed categorization task trying to seem as depressed as possible. It will not be obvious to understand how to do so; try to guess the mechanisms of the task. However, we recommend that you do not respond randomly as this will annul your task. In particular, if the response times and / or errors of the items are too low or too high, the task gets annulled.