REVIEW

A Meta-Analysis of Neuropsychological Effort Test Performance in Psychotic Disorders

Ivan Ruiz[1] · Ian M. Raugh[1] · Lisa A. Bartolomeo[1] · Gregory P. Strauss[1]

Abstract

Psychotic disorders are characterized by a generalized neurocognitive deficit (i.e., performance 1.5 SD below controls across neuropsychological domains with no specific profile of differential deficits). A motivational account of the generalized neurocognitive deficit has been proposed, which attributes poor neuropsychological testing performance to low effort. However, findings are inconsistent regarding effort test failure rate in individuals with psychotic disorders across studies (0–72%), and moderators are unclear, making it difficult to know whether the motivational explanation is viable. To address these issues, a meta-analysis was performed on data from 2205 individuals with psychotic disorders across 19 studies with 24 independent effects. Effort failure rate was examined along with moderators of effort test type, forensic status, IQ, positive symptoms, negative symptoms, diagnosis, age, gender, education, and antipsychotic use. The pooled weighted effort test failure rate was 18% across studies and there was a moderate pooled association between effort failure rate and global neurocognitive performance (r = .57). IQ and education significantly moderated failure rate. Collectively, these findings suggest that a nontrivial proportion of individuals with a psychotic disorder fail effort testing, and failure rate is associated with global neuropsychological impairment. However, given that effort tests are not immune to the effects of IQ in psychotic disorders, these results cannot attest to the viability of the motivational account of the generalized neurocognitive deficit. Furthermore, the significant moderating effect of IQ and education on effort test performance suggests that effort tests have questionable validity in this population and should be interpreted with caution.

Keywords Generalized neurocognitive deficit · Performance validity test · Psychotic disorders · Intellectual functioning · Effort

Neuropsychological impairment is a core feature of psychotic disorders that significantly predicts poor functional outcome (Bowie et al., 2008; Green et al., 2008; Keefe, Poe, Walker, Kang, & Harvey, 2006). Meta-analyses of neuropsychological test performance indicate that individuals with psychotic disorders (e.g., schizophrenia, schizoaffective disorder, psychosis NOS, bipolar with psychosis) perform approximately 1.5 standard deviation below healthy controls (Dickinson, Ramsey, & Gold, 2007; Fioravanti, Carlone, Vitale, Cinti, & Clare, 2005; Heinrichs & Zakzanis, 1998), with impairment across a range of cognitive domains and no distinctive pattern of differential deficits (Dickinson, 2008; Dickinson, Iannone, Wilk, & Gold, 2004; Dickinson, Ragland, Gold, & Gur, 2008; Reichenberg & Harvey, 2007). Such findings have led some to propose that psychotic disorders are characterized by a "generalized neurocognitive deficit" (Dickinson & Harvey, 2009; Dickinson et al., 2008).

Several accounts have been proposed to explain the generalized neurocognitive deficit. For example, studies support a role for abnormalities in grey- and white-matter, reduced signal integration across neural networks, neuropathology at the cellular level (e.g., NMDA receptor dysfunction, GABA interneurons), and abnormalities in "general systems" that impact the brain (e.g., inflammatory, metabolic, and oxidative stress processes; Dickinson & Harvey, 2009). These central nervous system and general systems theories have led to important advances in our understanding of the generalized neurocognitive deficit. However, they account for a limited proportion of variance in global neuropsychological test scores (1–34%) and have led to minimal treatment

breakthroughs (Gur, Turetsky, Bilker, & Gur, 1999; Sullivan, Shear, Lim, Zipursky, & Pfefferbaum, 1996; Zipursky, Lambe, Kapur, & Mikulis, 1998). Alternative approaches to understanding the generalized neurocognitive deficit may therefore be beneficial.

One possibility is that the generalized neurocognitive deficit in individuals with psychotic disorders results from problems with motivation that undermine the ability to allocate sufficient effort during testing. This proposal has appealing face validity for explaining the generalized neurocognitive deficit. If a significant proportion of individuals with psychotic disorders put forth insufficient effort on neuropsychological tests, this could explain why they display impairments across a wide range of cognitive domains with no specific profile of differential deficits.

There has been considerable interest in determining whether individuals with psychotic disorders put forth inadequate effort on neuropsychological tests. Insufficient effort is typically examined using standardized measures that are designed to look like difficult tests of cognitive ability but are in fact quite easy. These tests can be embedded within neuropsychological tests or batteries designed to quantify level and type of cognitive impairment (e.g., Reliable Digit Span; Iverson & Tulsky, 2003), or free-standing tests designed specifically for the purpose of indexing insufficient effort (e.g., Word Memory Test; Green, Iverson, & Allen, 1999). The key outcome from both free-standing and embedded effort tests is a binary score indicating whether adequate effort has been allocated (i.e., pass/fail). Instances where examinees fail effort tests typically result from either incentive to perform poorly (e.g., during litigation, disability determination) that causes feigned impairment (i.e., malingering; Mittenberg, Patton, Canyock, & Condit, 2002), or when the examinee is apathetic and lacks the intrinsic motivation to perform well (Konstantakopoulos et al., 2011).

The validity of free-standing and embedded effort testing is supported by studies indicating that neuropsychiatric patients with severe cognitive impairment and intellectual disability (i.e., IQ < 70) can obtain near perfect scores on these measures (Green, Rohling, Lees-Haley, & Allen, 2001; Rees, Tombaugh, Gansler, & Moczynski, 1998), suggesting that failure reflects motivation rather than cognitive impairment. However, there are some exceptions, such as severe dementia, where effort tests are not thought to be valid because patients fail them due to genuine cognitive deficits (Bianchini, Mathias, & Greve, 2001; Chafetz & Dufrene, 2014; Dean, Victor, Boone, & Arnold, 2008; Larrabee, 2014; Merten, Bossink, & Schmand, 2007; Singhal, Green, Ashaye, Shankar, & Gill, 2009; Smith et al., 2014; Teichner & Wagner, 2004; Tombaugh, 1996; Victor & Boone, 2007). Although effort testing is commonly used with individuals with psychotic disorders in clinical and forensic settings for a multitude of purposes that have important legal and health

ramifications, it is unclear whether effort tests are ideally suited for use in this population. It is therefore unclear whether these tests can also be used to make meaningful inferences about whether low effort contributes to the generalized neurocognitive deficit and whether they should be used in routine clinical practice.

A number of studies have used standardized embedded or free-standing tests to examine neuropsychological effort test failure rates in psychotic disorders. Results have been highly variable across studies, with failure rates ranging from 0 to 72% (see results section below). This is similar to the range of effort test failure rates found in other clinical groups with comparable or more severe cognitive deficits, including patients with traumatic brain injury at 10.7–59.1% (Armistead-Jehle, 2010; Armistead-Jehle & Buican, 2012; Lange, Sonal, Bhagwat, Anderson-Barnes, & French, 2012; Lippa, Agbayani, Hawes, Jokic, & Caroselli, 2015; Lippa, Lange, French, & Iverson, 2018; Nelson et al., 2010; Whitney, Shepard, Mariner, Mossbarger, & Herman, 2010; Whiteney, Shepard, Williams, Davis, & Adams, 2009), Parkinson's disease at 8–62.6% (Carter, Scott, Adams, & Linck, 2016), Huntington's disease at 18–70% (Sieck, Smith, Duff, Paulsen, & Beglinger, 2013), and across dementia disorders at 5–76% (Bortnik, Homer, & Bachman, 2013; Burton, Enright, O'Connell, Lanting, & Morgan, 2015; Davis, 2018; Duff et al., 2011; McGuire, Crawford, & Evans, 2019). Several factors may influence inconsistent failure rates across studies, including sample characteristics (e.g., symptoms, demographics, IQ), psychiatric hospitalization status (e.g., inpatient vs. outpatient), forensic status (e.g., litigating vs. not), the type of effort test used (i.e., freestanding vs. embedded), the cognitive domain measured by the test, and differences in the sensitivity and specificity of effort tests administered. To date, there has yet to be a systematic exploration of factors moderating effort test failure rates in psychotic disorders. However, there are some trends within the literature that suggest possible candidates for moderators.

The strongest candidate for a moderator appears to be negative symptoms, as several studies point to greater effort test failure rates among those with higher negative symptoms, particularly avolition (Avery, Startup, & Calabria, 2009; Gorissen, Sanz, & Schmand, 2005; Strauss, Morra, Sullivan, & Gold, 2015). This finding is consistent with the intuitive notion that motivational deficits lead individuals with psychotic disorders to put forth low effort and perform poorly on a range of neuropsychological tests. In contrast, associations with positive symptoms have generally been nonsignificant (Foussias et al., 2015; Moore et al., 2013; Morra, Gold, Sullivan, & Strauss, 2015; Strauss et al., 2015; Whearty, Allen, Lee, & Strauss, 2015).

Forensic status may be another relevant moderator. Studies have found that neuropsychiatric patients broadly defined, who are undergoing a legal case, are more likely to fail effort

testing. For example, up to 28% of individuals diagnosed with depression who were tested in a forensic context were found to produce invalid effort (Green et al., 2001; Mittenberg et al., 2002). Other studies have documented that the availability of secondary gains (e.g., applying for disability benefits, avoiding criminal persecution, etc.) is associated with malingering (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Rohling, Binder, & Langhinrichsen-Rohling, 1995). Using a heterogeneous sample, Paradis, Solomon, Owen, and Brooker (2013) found that the incidence of psychotic disorders was higher in individuals who failed an effort test compared to those who passed. Hence, it is possible that being involved in a civil forensic case may increase the likelihood of effort test failure in psychotic disorders. However, since the vast majority of studies examining effort failure rates to date have been conducted in a pure research environment, with non-litigating individuals who had little apparent incentive to feign impairment (i.e., malinger), forensic status seems unlikely to be a strong explanation for increased effort test failure rates in psychotic disorders.

It is unclear whether effort test type (i.e., free-standing vs. embedded) has a differential impact on failure rate in psychotic disorders. In the broader effort testing literature, there is some evidence that embedded measures have better classification accuracy (Miele, Gunner, Lynch, & McCaffrey, 2012). However, other studies warn against using embedded effort tests with cognitively impaired individuals because they may lead to false positives (Slick, Hopp, Strauss, & Thompson, 1997; Teichner & Wagner, 2004). It is unclear how effort test failure rates might differ as a function of whether the measure used is embedded or free-standing in psychotic disorders.

Demographic characteristics, such as age and education, are also known to impact effort test failure rates across a range of neuropsychiatric disorders (Rees et al., 1998; Tombaugh, 1996, 1997). Psychotic disorders have been associated with reduced personal educational attainment (Keefe, Eesley, & Poe, 2005) and accelerated aging (Kirkpatrick et al., 2008). However, it is unclear whether demographic factors might explain effort test failure rate in this population.

Finally, it has become clear that effort tests are not immune to the effects of genuine intellectual disability, and IQ may also be another relevant moderator. IQ has been associated with failure rates in several neuropsychiatric conditions (Chafetz & Dufrene, 2014; Dean et al., 2008; Larrabee, 2014; Smith et al., 2014; Victor & Boone, 2007), and a growing literature suggests that this may be true in psychotic disorders as well (Dean et al., 2008; Whearty et al., 2015). To depict this association, Table 1 presents a re-analysis of two studies conducted by our group, indicating that only participants with estimated Full-Scale IQ scores <80 failed neuropsychological effort tests more often (Morra et al., 2015; Strauss et al., 2015). Individuals with psychotic disorders with average and higher IQ scores did not fail more often. Thus, it is possible that these disorders could be among the

neuropsychiatric disorders with intellectual disabilities severe enough to produce elevated failure rates on effort tests due to genuine cognitive impairments.

Together, prior studies raise several important questions:

(1) What is the mean pooled effort test failure rate in individuals with psychotic disorders? Answering this question is of critical importance given the wide range of failure rates reported across studies (0–72%).

(2) Which variables significantly moderate effort test failure rates in individuals with psychotic disorders? The literature reviewed above identifies several plausible moderators, including: effort test type (embedded vs. free-standing), forensic status, estimated IQ, positive and negative symptom severity, and demographic variables. Answering this question has important implications for determining which factors contribute to effort test failure and whether effort tests are valid for use in clinical and research settings in this population.

(3) Is a low effort hypothesis a viable explanation for the generalized neurocognitive deficit? If such an account is viable, this would suggest that psychosocial, pharmacological, and cognitive rehabilitation approaches to treatment should target motivational processes in an attempt to enhance cognition. Few interventions for cognitive impairment have incorporated motivational elements and the infusion of such elements would reflect a novel approach with potential to improve cognition if low effort is a contributing factor (Velligan, Kern, & Gold, 2006).

The current study evaluated the aforementioned questions via a meta-analysis of data from 2205 individuals with psychotic disorders taken from 19 studies with 24 independent effects. It was hypothesized that: 1) a nontrivial proportion (i.e., >10%) of individuals with psychotic disorders would fail effort testing; 2) effort test failure would be moderated by IQ and negative symptoms, but not effort test type, forensic status, positive symptoms, or demographic variables; 3) global neuropsychological impairment would be predicted by effort test failure rate, which may be consistent with a motivational account of the generalized neurocognitive deficit.

Method

Protocol and Registration

This meta-analysis followed the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement guidelines (Liberati et al., 2009; Moher et al., 2009). The review protocol was pre-registered at PROSPERO International Prospective Register of Systematic Reviews (CRD42018086018) and can be accessed at: https://www.

Table 1 Reanalysis of the Effects of Estimated Full-Scale IQ on Effort Test Performance

Proportion of Each Estimated FSIQ Band Failing the RBANS &WMT

% Within FSIQ Group	FSIQ Group				Test Statistic, p value
	≤79	80–89	90–99	≥100	
RBANS Effort Index					
Fail (N = 31)	87.1%	9.7%	3.2%	0.0%	$X^2 = 20.49$
Pass (N = 299)	45.2%	21.1%	16.4%	17.4%	$p < 0.001$
Word Memory Test					
Fail (N = 7)	57.1%	42.9%	0.0%	0.0%	$X^2 = 8.79$
Pass (N = 38)	15.8%	29.0%	29.0%	26.0%	$p = 0.03$

FSIQ derived from Information, Block Design, Arithmetic, and Symbol Search subtests

crd.york.ac.uk/prospero/display_record.php?RecordID=86018.

Eligibility Criteria

The following inclusion criteria were implemented: (1) empirical articles published in a peer-reviewed journal, (2) use of at least one embedded or free-standing effort test, (3) inclusion of a psychosis group, and (4) available in English. Studies were excluded if they did not include sufficient data for effect size calculation (percent failure) and attempts to obtain missing data from the corresponding authors were unsuccessful. Studies that included multiple psychotic groups (i.e., independent effects) were included. For studies in which the same participants worked on multiple effort tests, a mean effect (percent failure) was calculated instead of arbitrarily selecting a failure rate from one of the effort tests reported. This method was considered necessary in order to only include independent effects.

Information Sources

A publication search was performed using the PsycINFO and PubMed databases between March 17th, 2018 and July 31st, 2018. Once relevant publications were identified, Google Scholar was used to manually search for eligible studies citing the already-identified publications that did not appear in the database search.

Search

The following key terms were entered in each database: "schizophrenia" OR "schizoaffective" OR "psychosis" OR "psychotic" OR **"psychotic disorder"** AND "neuropsychological test" OR "neuropsychological test of effort" OR "performance validity" OR "symptom validity" OR "malingering" OR "faking bad" OR "feigned impairment" OR "non-credible respon*" OR "Effort" OR "Motivation". The reference lists of relevant publications were searched

manually. Exclusion criteria were: duplicate articles, theses, dissertations, reviews and meta-analyses, book chapters, and poster presentations.

Study Selection

An initial search was conducted by using the specified search strategy. All titles were screened to identify relevant publications for full-text retrieval while abstracts were analyzed as needed. Full texts were retrieved and examined to confirm inclusion. For longitudinal studies, baseline scores were used. Studies that included measures that could have been used as embedded measures of effort (e.g., WAIS-IV Digit Span) but did not report percent effort failure were excluded.

Data Collection Process

A list of relevant variables was first developed by the first author, which was then revised after consulting with the last author. Corresponding authors from relevant publications with insufficient data to extract the main effect or moderator variables were contacted by the first author with a data request message. A follow-up email was sent approximately one month after the first email. To minimize experimenter bias, three coders independently extracted main effects and related data in which 18 effects were extracted consistently across all three coders and 4 effects were extracted consistently from two coders. An ICC = .83 was obtained, suggesting good reliability between coders. Coding discrepancies were resolved with discussion and consensus.

Data Items

The following variables were extracted from relevant publications: (1) study, (2) effort test, (3) effort test type (embedded vs. free-standing), (4) forensic status (forensic vs. non-forensic), (5) effort predicted by IQ (yes vs. no), (6) effort predicted by negative symptoms (yes vs. no), (7) sample size, (8)

%failure, (9) effort predicted by positive symptoms (yes vs. no), (10) IQ measure, (11) IQ *t*-score, (12) premorbid IQ measure, (13) premorbid *t*-score, (14) depression measure, (15) depression score (%), (16) positive symptom measure, (17) positive symptom score (%), (18) negative symptom measure, (19) negative symptom score (%), (20) %treated with antipsychotic medication, (21) Chlorpromazine equivalent (mean), (22) Chlorpromazine equivalent (*SD*), (23) Olanzapine equivalent (mean), (24) Olanzapine equivalent (*SD*), (25) %first generation antipsychotic only, (26) %second generation antipsychotic only, (27) %first and second antipsychotic generation, (28) age (mean), (29) age (*SD*), (30) personal education (mean), (31) personal education (*SD*), (32) parental education (mean), (33) parental education (*SD*), (34) %male, and (35) %White.

Risk of Bias in Individual Studies

The Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies was adapted to assess risk of bias for each publication included (National Heart, Lung, and Blood Institute, 2014). This measure was designed to assist reviewers in their appraisal of the internal validity of a study. Using this measure, each publication was evaluated on the following areas: 1) *Was the research question clearly defined?*, 2) *Was the study population clearly specified?*, 3) *Was the participation rate of eligible persons at least 50%?*, 4) *Were all the subjects recruited from the same or similar populations?*, 5) *Was a sample size justification, power description, or variance and effect estimates provided?*, 6) *Were the independent variables clearly defined, valid, reliable, and implemented consistently across all study participants?*, 7) *Was the exposure assessed more than once over time?*, 8) *Were the dependent variables clearly defined, valid, reliable, and implemented consistently across all study participants?*, 9) *Was the outcome assessor blind to the exposure status of participants?*, and 10) *Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure and outcome?* Each domain received a score of 0 (yes) or 1 (no; cannot be determined, or not reported). Publications received a risk bias total score (in which higher scores suggested higher risk bias), and this score was used in moderator analysis.

Summary Measures

The effect size of this meta-analysis was the percent (proportion) of effort test failure. Since this effect was not normally distributed (Shapiro-Wilk test: .86, *df* = 24, *p* = .003), effects were transformed to *logit* units. In this method, logit units are normally distributed, have a mean of zero

and a standard deviation of 1.83. All analyses were computed using logit units using the formulas,

$$ES = Log_c\ [p/(1-p)].SE = \sqrt{[1/(np) + 1/(n(1-p))]}.W$$
$$= 1/(SE^2) = np(1-p).$$

where *p* is the effort failure rate and *n* is the total number of participants for each effect. Final results were transformed back to percentages for an easier interpretation using the formulas,

$$p = e^x/(e^x + 1).\% = 100p.$$

where *x* is the *ES*.

Independent effects were extracted from each study. When multiple effects were derived from the same study (e.g., same group of participants participated in multiple effort tests), the multiple effects (in logit units) were averaged and included as an independent effect. This procedure was established a priori and considered necessary in order to include studies providing multiple effects from the same group. When single studies reported effort failure rates from independent samples, one effect was extracted and included from each sample.

Synthesis of Results

A mean pooled effect size was calculated using a maximum-likelihood random effects model, as random differences were expected to exceed heterogeneity attributed by subject-level sampling error alone (i.e., due to significant heterogeneity in symptom severity, patient characteristics, effort tests used, etc.). To examine the potential distortion of the transformed relative to untransformed effects, we also meta-analyzed untransformed effects to determine whether a significant discrepancy between the two sets of results was evident. We also meta-analyzed transformed effects using a fixed-effects model to explore whether significant discrepancies exist between models similarly to what previous meta-analyses have done (Martin et al., 2020; Sugawara et al., 2018). Heterogeneity between effects was tested using the *Q* statistic, which has a chi-square distribution with *k* − 1 degrees of freedom where *k* is the number of independent effects (Hedges & Olkin, 1985). I^2 was calculated as an estimate of the percentage of variance that is attributable to true variance rather than sampling error (Borenstein, Hedges, Higgins, & Rothstein, 2011; Higgins, Thompson, Deeks, & Altman, 2003). T^2 was also calculated as an estimate of true and absolute variance between effects (Borenstein et al., 2011). All statistical analyses were conducted in R (version 3.6.02019-04-26; Viechtbauer, 2010) using metafor package (version 2.1–0).

Risk of Bias across Studies

A fail-safe procedure was implemented to assess for publication bias (Rosenberg, 2005). In this method, a number of null effects, N_+, (i.e., effects at the estimated weighted failure rate) that represents a number of unpublished studies needed to move estimates to a non-significant difference is calculated (Rosenberg, 2005). When $N_+ = 5n + 10$, where n is the number of independent effects included in the meta-analysis, a fail-safe number is considered robust (Rosenberg, 2005). Moreover, the Rosenberg's fail-safe procedure is a weighted fail-safe calculation applicable to both fixed- and random-effects models (Rosenberg, 2005). A funnel plot was also created for a visual representation of the retrieved effects.

Additional Analyses

Univariate regression analyses were conducted to examine whether each of the predetermined moderator variables independently contributed to the heterogeneity in the pooled weighted effect. Categorical variables were dichotomized. The moderator variables considered for moderator analyses included the following: *effort test type*: 0 = embedded, 1 = freestanding; *forensic status*: 0 = non-forensic, 1 = forensic; *IQ* (*t*-scores); *premorbid IQ* (*t*-scores); *depression* (%); *negative symptoms* (%); *positive symptoms* (%); *disorganization* (%); *age* (years); *education* (years); *gender* (% male); and *medication doze* (chlorpromazine equivalent).

Studies that included data on positive symptoms reported using the Brief Psychiatric Rating Scale-Positive (Overall & Gorham, 1962), Positive and Negative Syndrome Scale-Positive (Kay, Fiszbein, & Opler, 1987), and Scale for the Assessment of Positive Symptoms (Andreasen, 1984), whereas studies that included data on negative symptoms utilized the Brief Negative Symptom Scale (Kirkpatrick et al., 2018; Kirkpatrick et al., 2011; Strauss & Gold, 2016; Strauss et al., 2012a; Strauss et al., 2012b; Strauss, Vertinski, Vogel, Ringdahl, & Allen, 2016), Scale for the Assessment of Negative Symptoms (Andreasen, 1982), Schedule for the Deficit Syndrome (Kirkpatrick, Buchanan, McKenney, Alphs, & Carpenter Jr., 1989), Positive and Negative Syndrome Scale-Negative (Kay et al., 1987), or Brief Psychiatric Rating Scale-Negative (Overall & Gorham, 1962). Since there is no single measure for assessing symptom severity in these dimensions of psychopathology in psychotic disorders, composite scores were calculated to put each symptom severity score on the same metric across scales by dividing the total symptom score observed by the maximum possible score for that scale or subscale. Similar methods have been used in past meta-analyses (Fusar-Poli et al., 2013; Llerena, Strauss, & Cohen, 2012; Wykes, Huddy, Cellard, McGurk, & Czobor, 2011).

To evaluate the association between neuropsychological effort and global cognitive performance, correlations reported in eligible publications between effort and global cognition were meta-analyzed using a Fisher's r to z transformation method (Fisher, 1921).

Results

Study Selection

An electronic search using the aforementioned keywords within the Title produced a total of 364 publications (PsycINFO = 188 studies; PubMed = 178). After duplicates were removed, a total of 228 unique publications were examined for eligibility. Out of those, irrelevant publications were excluded via title ($n = 164$), abstract ($n = 36$), or full-text ($n = 16$) review. Twelve eligible publications resulted from this search. Additionally, publications that have cited each of the relevant studies were found via Google Scholar and screened for eligibility ($n = 1073$) using the same inclusion criteria, producing seven eligible publications. A total of 19 studies were included and 24 independent effects were extracted from these studies (see Fig. 1 for a literature search flowchart).

Study Characteristics

Tables 2 and 3 reports study characteristics, including sample size, demographic variables, effort test used, effort test type, patient status, group type, and percent failure rate (raw and logit-transformed). In total, 19 unique articles providing $k = 24$ independent effects were included in this meta-analysis with a total sample size of 2205 psychotic disorder individuals ($M = 91.9$, $SD = 123.9$, Median = 58, range = 20–595). Out of the 16 samples that reported ethnicity, a total of 598 (42%) individuals were White. Out of the 22 samples that reported gender, 1301 (70%) were male. Out of the 22 samples that reported age, mean age was 39.3 ($SD = 10.4$). Eighteen samples included years of personal education, which averaged 11.8 ($SD = 2.9$). Data from 10 studies indicated that 93.5% of the patients were medicated with an antipsychotic. Three studies reported medication dose and four studies reported medication type. The effort test most frequently used was the TOMM (7 studies). The Reliable Digit Span was used by six studies. Finger Tapping, Word Memory Test, and RBANS-Effort Index were used by three studies. Dot Counting, CVLT Forced-Choice, Validity Indicator Profile Non-Verbal, Validity Indicator Profile-Verbal, Rey 15-Item, and Victoria Symptom Validity Test were each used by two studies. Other effort measures used in a single study included: Hiscock Forced-Choice, Digit Span Age Corrected, Rey Word Reading Test, Rey/Osterrieth Effort Equation, RAVLT Effort Equation, Recognition Memory Test,

Fig. 1 Flowchart of the Literature Search Process
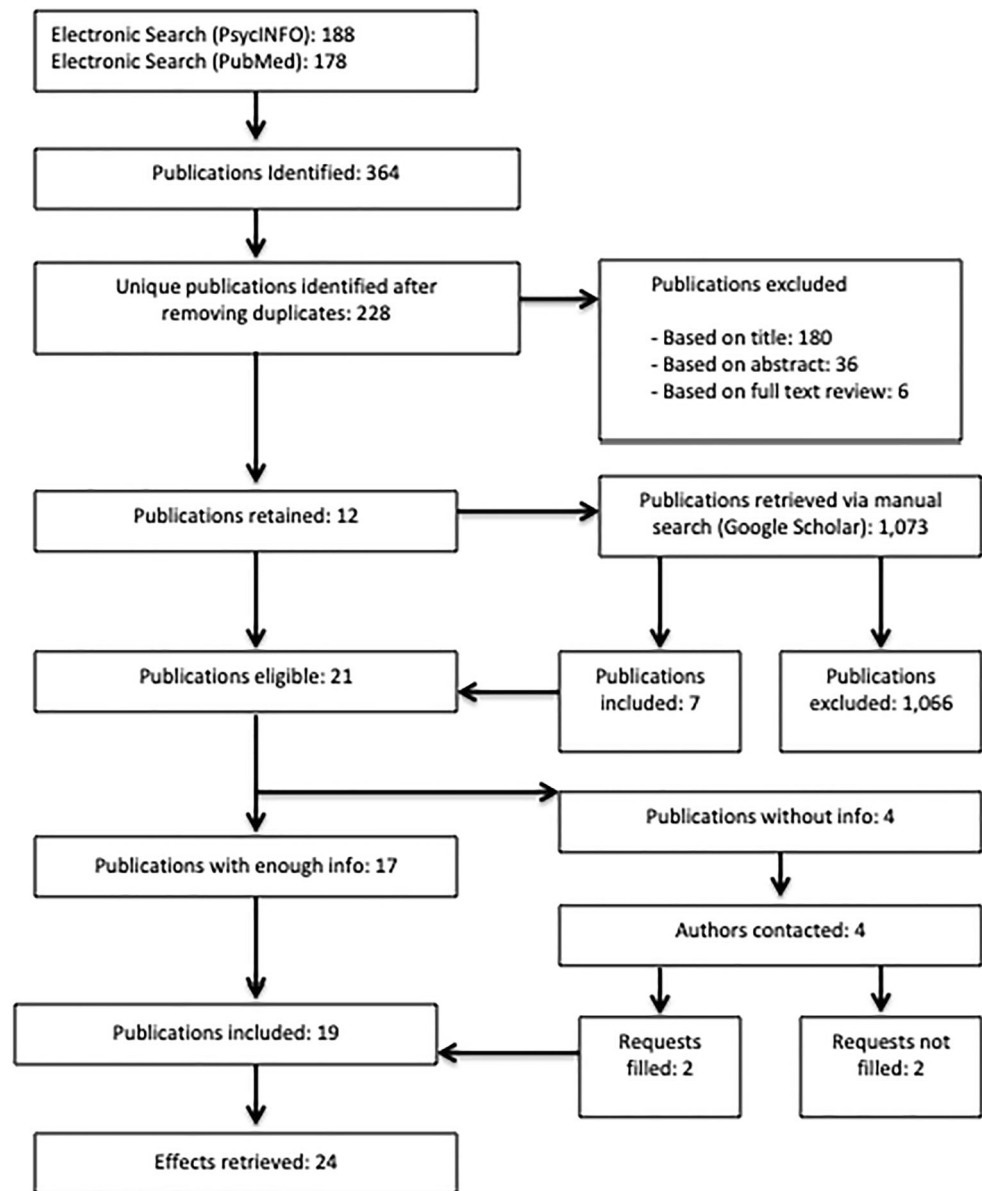
Sentence Repetition, Logical Memory Test Rarely Missed Items, Rey Complex Figure Test Recognition, and the 21-Item Test French Version.

Risk of Bias within Studies

The potential risk bias within studies was examined in multiple ways. First, all publications received a Quality Assessment Tool total score (0–10; higher scores = higher bias risk), and the majority of the studies obtained 3 or less points and none of these obtained >5 on this metric, suggesting adequate internal validity. Second, a moderator analysis using Quality Assessment Tool total score revealed that this metric was not a statistically significant moderator in effort failure ($R^2 = .15$, $\beta = .44$, $SE = .27$, $z = 1.66$, $p = .10$). Third, when

we excluded the four effects with the highest Quality Assessment Tool total score, results produced similar findings (effort failure = 15%, 95% CI: 10–22%, $Q = 154.15$, $p < .0001$). Therefore, risk of bias within studies was considered low.

Synthesis of Results

Percent failure rate across studies ranged from 0.0–72.0%. Across 24 effects from 19 studies, meta-analysis yielded a pooled weighted effort test failure rate of 18% (95% CI: 12–26%; See Fig. 2). We also meta-analyzed untransformed effects to examine the potential distribution impact of the logit-transformation procedure, and results revealed a similar pooled failure rate (20%; 95% CI: 13–27%). Exploratory

Table 2 Sample Characteristics of Studies Included in the Meta-Analysis

Study	N	Mean Age (years)	Mean Education (years)	Male (%)	White (%)
Back et al., 1996	30	34.9	13.2	66.7	73.3
Bayan, Nitch, Kinney, & Kaminetskaya, 2018	56	46.1	10.9	83.9	46.4
Bayan et al., 2018	595	41.7	11.0	72.1	38.2
Dean et al., 2008	20	40.2	11.8	45.0	40.0
Duncan, 2005 (1)	29	36.7	11.9	72.0	76.0
Duncan, 2005 (2)	21	36.5	11.8	62.0	71.0
Egeland et al., 2003	53	31.5	13.3	62.3	–
Foussias et al., 2015	69	38.0	–	70.0	–
Frederick, Crosby, & Wynkoop, 2000	60	–	–	–	–
Gierok, Dickson, & Cole, 2005 (1)	20	32.7	–	100.0	20.0
Gierok et al., 2005 (2)	20	41.0	–	100.0	75.0
Glassmire, Toofanian Ross, Kinney, & Nitch, 2016 (1)	88	43.2	11.3	79.8	6.0
Glassmire et al., 2016 (2)	56	–	–	–	–
Gorissen et al., 2005	64	55.9	11.6	71.9	–
Hunt, Root, & Bascetta, 2014	54	51.0	10.0	49.0	63.0
Moore et al., 2013	128	51.9	12.3	62.5	56.3
Morra et al., 2015	330	–	–	–	–
Pivovarova, Rosenfeld, Dole, Green, & Zapf, 2009	86	40.9	11.7	73.9	15.9
Raffard, Capdevielle, Boulenger, Gely-Nargeot, & Bayard, 2014	49	29.7	11.2	74.0	–
Schroeder & Marshall, 2011	104	35.7	12.9	69.2	–
Stevens et al., 2014	70	35.6	10.7	67.0	–
Strauss et al., 2015 (1)	97	42.1	12.5	69.1	57.7
Strauss et al., 2015 (2)	46	41.7	12.4	60.9	60.9
Whearty et al., 2015	60	39.8	11.9	53.3	48.3

analyses revealed a similar failure rate when implementing a more conservative approach for comparison based on a fixed-effects model (pooled failure rate = 22%; 95% CI: 20–24%). The weighted effort failure rate across studies using the logit method indicated that a nontrivial number of patients with a psychotic disorder failed an effort test. A test of heterogeneity indicated significant heterogeneity across the effects (Q = 243.0, $p < .001$). I^2 was 93 and T^2 was 1.12, suggesting that a significant amount of variance within the effects could be accounted for by true heterogeneity (Higgins et al., 2003) and not sampling error alone.

A sensitivity analysis was conducted to examine whether the pooled mean failure rate significantly changed after systematically removing each of the effects. Results indicated little change in the pooled effect size. Despite observing a drop in heterogeneity when the largest effect size was excluded from the analysis (Gorissen et al., 2005), the observed pooled effect was similar and heterogeneity was still large and significant (failure rate = 17%; 95% CI: 12–26%; Q = 178.4, $p < .001$).

Risk of Bias across Studies

Publication bias across studies was assessed using Fail-Safe N_+ test (Rosenberg, 2005). An N_+ of 787 suggests that this meta-analysis may not be affected by publication bias. N_+ is considered robust when it is larger than 5 k + 10, or 5*24 + 10 = 130 (Rosenberg, 2005). A funnel plot can be observed in Fig. 3, which displays asymmetry across the effects. When significant heterogeneity is expected and a sample includes a small number of effects, asymmetry is considered true heterogeneity due to moderator variables rather than publication bias (Sterne, Gavaghan, & Egger, 2000).

Moderator Analyses

Moderator analyses were conducted for each of the predetermined moderator variables using univariate regression analyses when there were at least 4 studies with sufficient data to conduct the moderator analysis. The following

Table 3 Details of Studies Included in Meta-Analysis

Study	Effort Test	Effort Test Type	Effort Failure (%)	Effort Failure (Logit)	Patient Status	Forensic Status	Affective Psychosis (%)	IQ Measure	Neurocognitive Battery	Depression Measure	Positive Symptom Measure	Negative Symptom Measure	% on Anti-psychotic	QAT
Back et al., 1996	R-15/RDC/HFC	FS	16.8	-1.6	both	non-forensic	0.0		—	—	—	—	96.7	3
Bayan et al., 2018(1)	CVLT-FC	EM	10.7	-2.12	inpatient	forensic	41.0		—	—	—	—	—	2
Bayan et al., 2018(2)	RBANS-EI	EM	15.6	-1.69	inpatient	forensic	34.9		RBANS	—	—	—	—	2
Dean et al., 2008	WAIS-III RDS and DS-ACSS, R-15, WRT, WRMT, RO-EE, RO-RAVLT-EE, DCT, FTT	Both	13.4	-2.9	outpatient	non-forensic	0.8	WAIS-III	—	—	—	—	—	2
Duncan, 2005(1)	TOMM	FS	14	-1.82	inpatient	forensic	40.4			—	—	—	—	2
Duncan, 2005(2)	TOMM	FS	0	-6.9	inpatient	forensic	33.0			—	—	—	—	2
Egeland et al., 2003	VSVT	FS	5	-2.94	both	non-forensic	0.0	WAIS-R: Similarities & Picture Completion	Selected: CVLT, RMT, RCFT, PASAT, WAIS-R DS	BPRS-Dep	PANSS-Pos	PANSS-Neg	92.5	2
Foussias et al., 2015	TOMM	FS	11.6	-2.03	outpatient	non-forensic	23.2		BACS	CDSS	SAPS	SANS	100	3
Frederick et al., 2000	VIP-NV	FS	45	-0.2	outpatient	forensic	6.7		—	—	—	—	—	4
Gierok et al., 2005(1)	TOMM	FS	30	-0.85	inpatient	forensic	—	WAIS-III: Vocabulary & Block Design	—	—	—	—	—	5
Gierok et al., 2005(2)	TOMM	FS	5	-2.94	inpatient	Non-forensic	—	WAIS-III: Vocabulary & Block Design	—	—	—	—	—	5
Glassmire et al., 2016(1)	RDS[a]	EM	33.3	-0.69	inpatient	non-forensic	38.1	WAIS-IV	—	—	—	—	—	3
Glassmire et al., 2016(2)	RDS[a]	EM	32.1	-0.75	inpatient	forensic	13.5	WAIS-IV	—	—	—	—	—	3
Gorissen et al., 2005	WMT	FS	72	0.94	both	non-forensic	0.0		Selected: WAIS or WAIS-III DS, Stroop Task, RAVLT or CVLT, WMS-R or WMS-III VR, SF	—	PANSS-Pos	PANSS-Neg	—	4
Hunt et al., 2014	TOMM/VIP-V/VIP-NV	FS	57	0.35	outpatient	non-forensic	37.0		—	—	—	—	—	2
Moore et al., 2013	RBANS-EI	EM	23	-1.23	outpatient	non-forensic	26.0		RBANS	CDSS	PANSS-Pos	PANSS-Neg	50	3
Morra et al., 2015	RBANS-EI	EM	9.4	-2.27	both	non-forensic	9.7	WASI-III	RBANS	—	SAPS	SDS	—	3
Pivovarova et al., 2009	TOMM/VIP-V	FS	13	-1.9	outpatient	non-forensic	—		—	—	—	—	—	3
Raffard et al., 2014	f-21 IT	FS	0	-6.9	outpatient	non-forensic	0.0		—	—	PANSS-Pos	PANSS-Neg	100	3
Schroeder & Marshall, 2011	RDS, SR, CVLT-FC, LM-RMI, FTT, DCT, RCFT-R	FS	4.2	-3.12	both	non-forensic	11.5	WAIS-III	Selected: WMS-III LM & VR, RCFT,	—	—	—	—	3

Table 3 (continued)

Study	Effort Test	Effort Test Type	Effort Failure (%)	Effort Failure (Logit)	Patient Status	Forensic Status	Affective Psychosis (%)	IQ Measure	Neurocognitive Battery	Depression Measure	Positive Symptom Measure	Negative Symptom Measure	% on Anti-psychotic	QAT
Stevens et al., 2014	WMT	FS	26	−1.05	inpatient	non-forensic	0.0	–	CVLT-II, D-KEFS (VF, DF, C-WI, ST) Selected: TAP, WMT, VVM, WCST-CV, TMT-A & B, WMS-R DS & BT, DSS	–	PANSS-Pos	PANSS-Neg	100	3
Strauss et al., 2015(1)	VSVT	FS	1	−4.6	outpatient	non-forensic	–	WASI	MCCB	BPRS-Dep	BPRS-Pos	SANS	100	2
Strauss et al., 2015(2)	WMT	FS	15.2	−1.72	outpatient	non-forensic	–	–	MCCB	BPRS-Dep	BPRS-Pos	SANS	–	2
Whearty et al., 2015	RDS/FTT	EM	31.9	−0.76	outpatient	non-forensic	21.7	WAIS-III	Selected: FTT, CVLT, BFLT-E, WCST, TMT A & B, WMS-III DS & SS, PPT	–	SAPS	SANS	91.7	3

Effort Test: *TOMM* Test of Memory Malingering, *VSVT* Victoria Symptom Validity Test, *WMT* Word Memory Test, *RCFT-R* Rey Complex Figure Task-Recognition, *R-15* Rey 15-Item Memory Test, *RDC* Rey Dot Counting, *HFC* Hiscock Forced-Choice Method, *CVLT-FC* California Verbal Learning Test-Forced-Choice, *RBANS-EI* Repeatable Battery for the Assessment of Neuropsychological Status-Effort Index, *RDS* Reliable Digit Span, *ACSS* Age Corrected Scale Score, *WRT* Rey Word Recognition Test, *WRMT* Warrington Recognition Memory Test-Words, *RO-EE* Rey-Osterreith Effort Equation, *RO-RAVLT-EE* Rey-Osterrieth Rey Auditory Verbal Learning Test-Effort Equation, *DCT* Dot Counting Test, *FTT* Finger Tapping Test, *f-21 IT* French Version of the 21-Item Verbal Memory Task, *SR* Sentence Repetition (raw score). Test type: *FS* Freestanding, *EM* Embedded. Neurocognitive Batteries/Measures: *RBANS* Repeatable Battery for the Assessment of Neuropsychological Status, *BACS* Brief Assessment of Cognition in Schizophrenia, *CVLT* California Verbal Learning Test, *RMT* Recognition Memory Test, *RCFT* Rey Complex Figure Test, *PASAT* Paced Auditory Serial Addition, *VIP-V/NV* Validity Indicator Profile-Verbal/Non-verbal, *VR* Visual Reproduction, *LM* Logical Memory, *SF* Semantic Fluency, *D-KEFS* Delkus-Kaplan Executive Function Systems, *VF* Verbal Fluency, *DF* Design Fluency, *C-WI* Color-Word Interference, *ST* Sorting Test, *TAP* Test for Attentional Performance (Sustained Attention and Alertness), *VVM* Visual and Verbal Memory, *WCST-CV* Wisconsin Card Sorting Test-Computer Version, *TMT* Trail Making Test, *BT* Block Tapping, *DSS* Digit Symbol Substitution, *SS* Spatial Span, *BFLT-E* Biber Figure Learning Test-Extended, *PPT* Purdue Pegboard Task. *QAT* Quality Assessment Tool

a Effort assessed by the Reliable Digit Span method using the standard cutoff score (≤6)

moderator variables were not tested, as the data extracted from the included studies was insufficient to run moderator analyses on these variables: premorbid IQ, depression, medication dose (chlorpromazine equivalent), and disorganization.

As shown in Table 4, the following moderators were non-significant: patient status, forensic status, negative symptoms, positive symptoms, age, gender, and % prescribed an antipsychotic. However, IQ was a significant moderator, such that patients with lower IQ were more likely to fail effort testing. This suggests that lower intellectual functioning is associated with higher effort test failure rate in individuals with psychotic disorders. Similarly, personal education was a significant moderator, suggesting that lower education is associated with higher effort test failure rate. A sensitivity analysis indicated that IQ and education remained significant moderators even when each sample was systematically removed from the analysis. Exploratory analyses examined whether proportion of affective (schizophrenia, psychotic disorder, psychosis NOS) versus nonaffective psychosis (schizoaffective, bipolar with psychosis, depression with psychosis) or percent medicated independently moderated failure rates. Results revealed that these moderator variables did not achieve statistical significance.

Does Effort Predict the Generalized Neurocognitive Deficit?

An analysis was conducted to examine whether effort test failure rate is associated with global neuropsychological functioning to evaluate the motivational account of the generalized neurocognitive deficit. Correlations between effort test failure rate and global neurocognitive performance from five independent effects were meta-analyzed using the Fisher's r to z method. Results revealed a significant association between effort and global cognition ($r = .57$; 95% CI = .44–.70; $p < .0001$), explaining a significant amount of variance ($R^2 = .32$). This correlation was significant even when Moore et al. (2013) was removed from the analysis, given that global neurocognition and effort were assessed using the same measure (RBANS and RBANS Effort Index) in this study ($r = .49$; 95% CI = .36–.61; $p < .0001$; $R^2 = .24$). This result should be interpreted with caution given that IQ was found to be a significant moderator and effort test performance may therefore reflect genuine cognitive impairment rather than true motivational deficits.

Discussion

Failure Rate

The first aim of this meta-analysis was to quantitatively synthesize the failure rate of effort test performance in psychotic disorders. Findings from 2205 individuals with psychotic disorders taken from 19 published studies with 24 independent effects indicated a pooled failure rate of 18%. This pooled failure rate was similar when using transformed and untransformed effects and fixed versus random models. Furthermore, when only effects derived from non-forensic or non-litigating samples were meta-analyzed ($k = 17$), a pooled mean failure rate of 16% was obtained, suggesting similar failure rates when patients do not have clear incentives for feigning impairment. These findings suggest that a non-trivial proportion of individuals with psychotic disorders fail effort testing and malingering is unlikely to account for the majority of effort test failures that are observed in psychotic disorders.
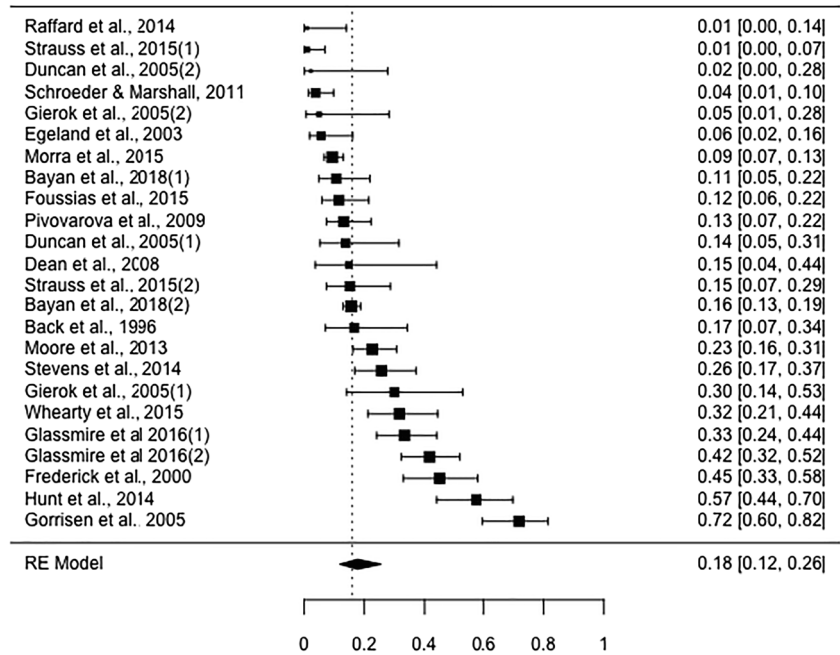
Significant Moderators of Failure Rate

To address the reasons why individuals with psychotic disorders fail effort testing, the second aim of the study was to evaluate moderators of failure rate. IQ and personal education were found to be significant moderator variables, explaining 91% and 32% of the variance, respectively.

Evidence that IQ moderates effort test failure rates does not lead to a clear interpretation. It is possible that low effort resulting from a motivational problem leads people with psychotic disorders to perform poorly on IQ or other cognitive tests or that low IQ or genuine cognitive impairments leads them to fail effort tests because these are simply other cognitive tests that people with genuine impairment perform poorly on. Alternatively, a combination of both genuine cognitive impairment and low motivation may lead to effort test failure in psychotic disorders.

Unfortunately, moderation analyses alone cannot clarify which of these explanations is most viable. One way that these competing explanations for effort test failure has been addressed in the literature is by calculating an "easy-hard" difference score on effort tests that allow for such comparisons (Green, Montijo, & Brockhaus, 2011; Howe, Anderson, Kaufman, Sachs, & Loring, 2007). Using such difference metrics, dementia patients with genuine profound cognitive impairments have been found to have large difference scores (i.e., they perform much worse on the hard than easy tests), whereas patients with low effort tend to perform equivalently on the easy and hard scores (i.e., they are lower on both).

Due to our inability to obtain data on easy and hard effort sub-test conditions from the studies included in this meta-analysis, we were unable to directly examine whether primarily profound cognitive impairments or primarily low effort moderates effort test failure rates in psychotic disorders meta-analytically. However, we were able to evaluate data from our two-experiment paper that used the VSVT and WMT in schizophrenia patients and healthy controls (Strauss et al., 2015) to draw some preliminary conclusions. On the VSVT, the control group demonstrated negligible differences

Fig. 2 Forest Plot. *Note.* Forest
plot of 24 independent effects
extracted from 19 publications.
Each effect includes first author,
publication year, proportion
(failure rate), and 95% confidence
interval. Mean pooled proportion
is located at the bottom

Raffard et al., 2014	0.01 [0.00, 0.14]
Strauss et al., 2015(1)	0.01 [0.00, 0.07]
Duncan et al., 2005(2)	0.02 [0.00, 0.28]
Schroeder & Marshall, 2011	0.04 [0.01, 0.10]
Gierok et al., 2005(2)	0.05 [0.01, 0.28]
Egeland et al., 2003	0.06 [0.02, 0.16]
Morra et al., 2015	0.09 [0.07, 0.13]
Bayan et al., 2018(1)	0.11 [0.05, 0.22]
Foussias et al., 2015	0.12 [0.06, 0.22]
Pivovarova et al., 2009	0.13 [0.07, 0.22]
Duncan et al., 2005(1)	0.14 [0.05, 0.31]
Dean et al., 2008	0.15 [0.04, 0.44]
Strauss et al., 2015(2)	0.15 [0.07, 0.29]
Bayan et al., 2018(2)	0.16 [0.13, 0.19]
Back et al., 1996	0.17 [0.07, 0.34]
Moore et al., 2013	0.23 [0.16, 0.31]
Stevens et al., 2014	0.26 [0.17, 0.37]
Gierok et al., 2005(1)	0.30 [0.14, 0.53]
Whearty et al., 2015	0.32 [0.21, 0.44]
Glassmire et al 2016(1)	0.33 [0.24, 0.44]
Glassmire et al 2016(2)	0.42 [0.32, 0.52]
Frederick et al., 2000	0.45 [0.33, 0.58]
Hunt et al., 2014	0.57 [0.44, 0.70]
Gorrisen et al., 2005	0.72 [0.60, 0.82]
RE Model	0.18 [0.12, 0.26]

between easy and hard conditions (24 vs. 23.7), whereas the psychotic disorder group demonstrated a larger difference (23.6 vs. 21.9). However, no subjects failed the VSVT in either sample, making it a less than ideal comparison for this purpose. On the WMT, which 15.2% of individuals with psychotic disorders failed, the differences between easy (mean of Immediate Recognition, Delayed Recognition, and Consistency) and hard (mean of Multiple Choice, Paired Associates, and Free Recall) conditions were even more pronounced in psychotic disorders (93.4% easy vs. 56.9% Hard). In comparison, controls scored 98.7% on easy and 87.2% on hard. This magnitude of difference in psychotic disorders WMT performance (36.5%) is comparable to that observed in cases with severe dementia (Green et al., 2011), consistent with the notion that individuals with psychotic disorders fail effort tests due to genuine cognitive impairment.
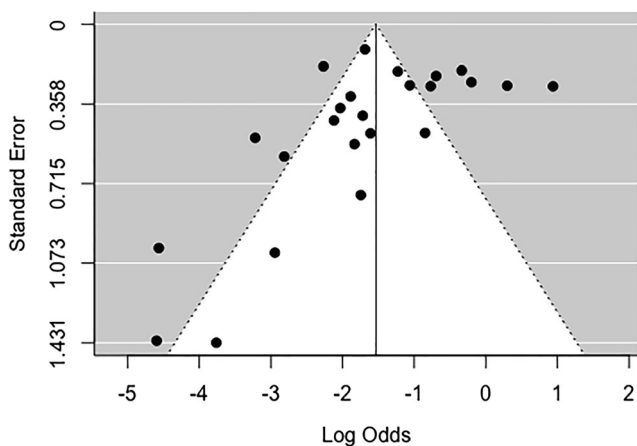

Fig. 3 Funnel Plot. *Note.* Funnel plot illustrating log odds (x-axis) by standard error (y-axis)

However, these group level results may be misleading. To further explore this "profile analysis", we evaluated individuals with psychotic disorders who failed versus passed in our WMT sample. When the psychosis sample who failed the WMT (Easy = 80%, Hard = 31%, = 49% difference score) are separated out from those who passed (Easy = 96%, Hard = 62%, = 34% difference score), it becomes clear that

Table 4   Moderator Analyses

Moderator	Logit Method				
	R²	β	SE	z	p value
Effort Test Type	.01	.10	.57	.18	.85
Embedded (7)	–	–	–	–	–
Free-standing (16)	–	–	–	–	–
Patient Status	.01	.02	.45	.05	.96
Inpatient (9)	–	–	–	–	–
Outpatient (10)	–	–	–	–	–
Forensic Status	.02	.32	.52	.61	.54
Forensic (7)	–	–	–	–	–
Non-forensic (17)	–	–	–	–	–
Estimated IQ (12)	**.91**	**−.20**	**.03**	**−6.01**	**<.001**
Negative Symptoms (9)	<.01	.01	.07	.22	.83
Positive Symptoms (7)	<.01	.01	.07	−.03	.98
Age (21)	.08	.06	.05	1.31	.19
Education (18)	**.32**	**−.66**	**.29**	**−2.30**	**.02**
Gender (%male) (22)	.02	−.01	.02	−.52	.60
% Medicated (9)	.20	.17	.13	1.36	.17
% Affective Psychosis (19)	.01	.06	1.81	.03	.97

the sample that fail perform very poorly on both easy and hard subtests and have an even greater difference score than the patients who pass. This profile analysis strongly speaks against a pure motivational account or combined motivational and genuine cognitive impairment account of why individuals with psychotic disorders fail effort test. One would expect a smaller, not larger, difference score in the patients who failed compared to those who pass if a motivational account was supported. However, the opposite was true, suggesting that individuals with psychotic disorders fail effort tests due to genuinely severe cognitive impairments.

To further explore the question of whether people with psychotic disorders fail effort tests due to low IQ alone or whether a combination of low IQ and motivational deficits contribute, we conducted supplemental mediation analyses on three of the studies in the meta-analysis that came from our lab (Morra et al., 2015; Strauss et al., 2015; Whearty et al., 2015; see Supplemental Materials for full details). Three mediation models were constructed to determine whether: 1) IQ mediated the association between global neurocognitive impairment and effort test failure, 2) motivational symptoms (avolition measured via a negative symptom clinical rating scale) mediated the association between global neurocognitive impairment and effort test failure, and 3) the combination of motivational symptoms and IQ mediated the association between global neurocognitive impairment and effort test failure. Results indicated that IQ significantly mediated the association between global neurocognitive impairment and effort test failure. However, motivational symptoms were not a significant mediator of the association between global neurocognitive impairment and effort test failure. The combined mediation of both IQ and motivational symptoms was significant for 1/3 studies and the significant model was driven by IQ. When coupled with the supplemental profile analyses (easy-hard discrepancies), these findings suggest that effort test failure in psychosis samples primarily results from low IQ or genuine cognitive impairments, rather than low motivation or a combination of low motivation and cognitive impairment. Thus, individuals with psychotic disorders may fail effort tests because they have genuine cognitive impairments and effort tests are simply another type of cognitive test that they perform poorly on due to genuine intellectual deficits.

Nonsignificant Moderators of Failure Rate

The observation that other moderators were nonsignificant is also noteworthy. A number of studies have found that greater negative symptom severity is associated with increased likelihood of failing effort testing (Avery et al., 2009; Foussias et al., 2015; Gorissen et al., 2005; Morra et al., 2015; Strauss et al., 2015), leading us to hypothesize that negative symptoms would be a significant moderator. The null result is therefore surprising. However, the finding should be interpreted with caution given that the majority of past studies used negative symptom scales that do not measure the construct per modern conceptualizations using first-generation negative symptom measures like the Brief Psychiatric Rating Scale (BPRS: Overall & Gorham, 1962) and Positive and Negative Syndrome Scale (PANSS: Kay et al., 1987). Associations between effort and negative symptoms have been stronger in studies using second-generation negative symptom rating scales, such as the Brief Negative Symptom Scale (BNSS: Kirkpatrick et al., 2011; Kirkpatrick et al., 2018; Strauss & Chapman, 2018; Strauss & Gold, 2016; Strauss et al., 2012a; Strauss et al., 2012b; Strauss et al., 2016; Strauss et al., 2018a; Strauss et al., 2018b). It is not surprising that first-generation scales would show little relation to effort. These scales measure the more expressive aspects of negative symptom pathology (e.g., alogia, blunted affect), and neglect the apathetic dimension (avolition, anhedonia, asociality) that should be more theoretically related to effort. Thus, it is unclear whether the nonsignificant negative symptom effect is reliable. Future studies should utilize the most conceptually updated measures (BNSS and CAINS; Kring, Gur, Blanchard, Horan, & Reise, 2013; Kirkpatrick et al., 2011) to test the role of negative symptoms in low effort most adequately.

Furthermore, the nonsignificant effects observed for other moderators are also meaningful. It is easy to imagine how delusions and hallucinations could interfere with cognitive testing, for example, leading patients to perform poorly and fail effort tests if distracted by internal or external stimuli. However, both positive symptoms and inpatient status (where positive symptoms are most pronounced) were not significant moderators, suggesting that the more florid aspects of psychosis psychopathology are contributing minimally to failure rates. Furthermore, other demographic factors that have been found to be associated with poor cognitive test performance (e.g., age) in other disorders were not significant moderators.

Is the Motivational Account of the Generalized Neurocognitive Deficit Viable?

The third aim of the study was to determine whether a motivational account of the generalized neurocognitive deficit might be viable. To examine this question, correlations between effort and global cognitive functioning were meta-analyzed from five independent effects. Results indicated that low effort accounted for 31% of the variance in global neurocognitive functioning. At first glance, these findings appear to support the motivational account of the generalized neurocognitive deficit. However, caution must be applied when interpreting this result given that IQ also moderated effort test failure. It is not clear whether low effort is a viable explanation for the generalized neurocognitive deficit because

effort tests may lack validity in the psychotic disorder population. Simply put, the finding that effort test failure predicts global cognitive test scores may simply reflect that poor performance on one cognitive test is associated with poor performance on another cognitive test. This correlation may have little to do with effort, but simply indicate that multiple cognitive tests are correlated with each other because effort tests have questionable validity in the psychotic population.

To evaluate this possibility directly, we conducted further supplemental mediation analyses on the data from the three studies previously published by our group (Morra et al., 2015; Strauss et al., 2015; Wheraty et al., 2015; see Supplemental Materials). Mediation models examined whether: 1) IQ mediated the link between effort and global neurocognitive impairment and 2) effort mediated the link between IQ and global cognitive impairment (see Supplemental Materials). Findings indicated that IQ was either a partial or full mediator between effort and global cognitive impairment in all 3 studies, whereas the model evaluating whether effort mediated the association between IQ and global cognitive impairment was significant in 1/3 studies and nonsignificant at a trend level in 2/3 studies. These findings suggest that the motivational account of the generalized neurocognitive deficit may be plausible. However, it is hard to have much confidence in such conclusions since they are drawn from neuropsychological effort tests. These measures may not be ideal to test the viability of the motivational account because their failure rates are heavily influenced by intellectual ability.

To accurately evaluate the question of whether low effort accounts for the generalized neurocognitive deficit in psychotic disorders, alternative approaches to assessing effort are needed, such as those developed in cognitive neuroscience (Kool, McGuire, Rosen, & Botvinick, 2010; Westbrook, Kester, & Braver, 2013). Using such tests, there is some evidence that poor cognition is associated with deficits in effort-cost computation, which are driven by motivational symptoms (Bismark et al., 2018; Gold et al., 2013). Cognitive neuroscience-based effort tests could be adapted for clinical purposes, standardized, normed, and tested to see if they have clinical utility for detecting the type of low effort that results from motivational impairments that are expected to contribute to the generalized neurocognitive deficit in psychotic disorders.

Limitations

Certain limitations should be noted. First, studies included in this meta-analysis utilized several types of embedded and free-standing tests that assessed effort via performance on specific cognitive domains (e.g., memory, attention, visual perception) or behaviors (e.g., finger tapping). Although these tests theoretically measure the same construct (i.e., effort), distinct levels of specificity and sensitivity across measures

might impact performance differently depending on the patient's characteristics. Prospective meta-analyses on specific effort tests would be informative. Second, results should be interpreted with the caveat that only a limited number of effects were included for certain moderator analyses. Additionally, a lack of data on premorbid IQ, disorganization, medication type and dose, and depression prevented the examination of moderating effects from these variables. Third, the approach taken to standardize positive and negative symptoms severity across measures is not ideal. However, in the absence of a universal measure to assess positive and negative symptoms severity, this approach was considered necessary to examine moderation effects of symptom severity on effort test failure rate. Relatedly, we were unable to examine the unique moderator effects of the most relevant negative symptom dimension (motivation/pleasure) due to the lack of available data on this dimension apart from total negative symptoms scores. Future studies may be able to address this question using more conceptually up-to-date rating scales (BNSS, CAINS). Lastly, we were unable to address the effects of illness stage due to insufficient number of studies in the first versus multi-episode samples.

Conclusions and Implications

In summary, across a heterogeneous sample of studies, patients with psychotic disorders are at higher chances of failing standard effort tests. Effort test failure is a significant predictor of global neuropsychological impairment. However, this association should not be taken as strong support for the motivational account of the generalized neurocognitive deficit given that effort test failure rates are influenced by genuine cognitive impairments in psychotic disorders. Thus, there is a need for new effort tests to be developed that are specifically designed to accommodate the cognitive capacity limitations of this population.

Our findings have important implications for the use of effort tests in clinical practice. Effort tests are now widely used to determine performance validity during neuropsychological evaluations. It is important for these tests to be valid since they play an important role for informing legal decisions pertaining to possible external benefits (e.g., disability status, compensation, etc.) or criminal sentencing. However, the moderation and supplemental mediation analyses reported in this study suggest that effort tests are not immune to the effects of low IQ in this population. Individuals with psychotic disorders appear to fail effort tests because they have genuine cognitive deficits and effort tests are simply cognitive tests that are not immune to true cognitive impairment or low IQ. Similar findings have been reported in other populations with more severe cognitive impairment, such as Alzheimer's dementia (Bianchini et al., 2001; Merten et al., 2007; Singhal et al., 2009; Teichner & Wagner, 2004; Tombaugh, 1996), and led

to the conclusion that there are "limits to effort testing" and the populations in which they should be used (Green et al., 2011; Merten et al., 2007). Our results suggest that psychotic disorders should be added to the list of disorders with severe cognitive impairment for which there are known to be "limits to effort testing." In situations where such tests must be used in individuals with psychotic disorders, it may be beneficial to administer multiple effort measures and engage in a multi-faceted decision-making processes using as many sources of information as are available, for example, clinical interview, behavioral observations, referral information, whether they are from a non-credible risk group, and effort test performance (Lezack, Howieson, Bigler, & Tranel, 2012; Rickards, Cranston, Touradji, & Bechtold, 2018; Slick, Sherman, & Iverson, 1999).

References

Andreasen, N. (1982). The scale for the assessment of negative symptoms (SANS). Iowa City, Iowa: The University of Iowa.

Andreasen, N. (1984). Scale for the assessment of positive symptoms (SAPS). Iowa City, University of Iowa:

Armistead-Jehle, P. (2010). Symptom validity test performance in U.S. veterans referred for evaluation of mild TBI. *Applied Neuropsychology*, 17(1), 52–59. https://doi.org/10.1080/09084280903526182

Armistead-Jehle, P., & Buican, B. (2012). Evaluation context and symptom validity test performance in a U.S. military sample. *Archive of Clinical Neuropsychology*, 27(8), 828–839. https://doi.org/10.1093/arclin/acs086

Avery, R., Startup, M., & Calabria, K. (2009). The role of effort, cognitive expectancy appraisals and coping style in the maintenance of the negative symptoms of schizophrenia (comparative study). *Psychiatry Research*, 167(1–2), 36–46. https://doi.org/10.1016/j.psychres.2008.04.016

Back, C., Boone, K. B., Edwards, C., Parks, C., Burgoyne, K., & Silver, B. (1996). The performance of schizophrenics on three cognitive tests of malingering, Rey 15-item memory test, Rey dot counting, and Hiscock forced-choice method. *Assessment*, 3(4), 449–457. https://doi.org/10.1177/107319119600300411

Bayan, S. M., Nitch, S., Kinney, D., & Kaminetskaya, M. (2018). Exploratory analysis of CVLT- II forced-choice recognition and RBANS EI base rates for forensically-committed, psychotic-disordered inpatients: Implications on clinical decision-making. *Applied Neuropsychology: Adult*, 25(1), 71–81. https://doi.org/10.1080/23279095.2016.1247092

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society*, 11(3), 215–227. https://doi.org/10.1017/S1355617705050277

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist*, 15(1), 19–45. https://doi.org/10.1076/clin.15.1.19.1907

Bismark, A. W., Thomas, M. L., Tarasenko, M., Shiluk, A. L., Rackelmann, S. Y., Young, J. W., & Light, G. A. (2018). Relationship between effortful motivation and neurocognition in schizophrenia. *Schizophrenia Research*, 193, 69–76. https://doi.org/10.1016/j.schres.2017.06.042

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). Introduction to meta- analysis. John Wiley & Sons.

Bortnik, K. E., Homer, M. D., & Bachman, D. L. (2013). Performance on standard indexes of effort among patients with dementia. *Applied Neuropsychology: Adult*, 20(4), 233–242. https://doi.org/10.1080/09084282.2012.695757

Bowie, C. R., Leung, W. W., Reichenberg, A., McClure, M. M., Patterson, T. L., Heaton, R. K., & Harvey, P. D. (2008). Predicting schizophrenia patients' real-world behavior with specific neuropsychological and functional capacity measures. *Biological Psychiatry*, 63(5), 505–511. https://doi.org/10.1016/J.Biopsych.2007.05.022

Burton, R. L., Enright, J., O'Connell, M. E., Lanting, S., & Morgan, D. (2015). RBANS embedded measures of suboptimal effort in dementia: Effort scale has a lower failure rate than the effort index. *Archives of Clinical Neuropsychology*, 30(1), 1–6. https://doi.org/10.1093/arclin/acu070

Carter, K. R., Scott, J. G., Adams, R. L., & Linck, J. (2016). Base rate comparison of suboptimal scores on the RBANS effort scale and effort index in parkinson's disease. *The Clinical Neuropsychologist*, 30(7), 1118–1125. https://doi.org/10.1080/13854046.2016.1206145

Chafetz, M., & Dufrene, M. (2014). Malingering-by-proxy: Need for child protection and guidance for reporting. *Child Abuse & Neglect*, 38(11), 1755–1765. https://doi.org/10.1016/j.chiabu.2014.08.015

Davis, J. J. (2018). Performance validity in older adults: Observed versus predicted false positive rates in relation to number of tests administered. *Journal of Clinical and Experimental Neuropsychology*, 40(10), 1013–1021. https://doi.org/10.1080/13803395.2018.1472221

Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist*, 22(4), 705–722. https://doi.org/10.1080/13854040701440493

Dickinson, D. (2008). Digit symbol coding and general cognitive ability in schizophrenia: Worth another look? *The British Journal of Psychiatry*, 193(5), 354–356. https://doi.org/10.1192/bjp.bp.108.049387

Dickinson, D., & Harvey, P. D. (2009). Systemic hypotheses for generalized cognitive bDeficits in schizophrenia: A new take on an old problem. *Schizophrenia Bulletin*, 35(2), 403–414. https://doi.org/10.1093/Schbul/Sbn097

Dickinson, D., Iannone, V. N., Wilk, C. M., & Gold, J. M. (2004). General and specific cognitive deficits in schizophrenia. *Biological Psychiatry*, 55(8), 826–833. https://doi.org/10.1016/j.biopsych.2003.12.010

Dickinson, D., Ragland, J. D., Gold, J. M., & Gur, R. C. (2008). General and specific cognitive deficits in schizophrenia: Goliath defeats David? *Biological Psychiatry*, 64(9), 823–827. https://doi.org/10.1016/j.biopsych.2008.04.005

Dickinson, D., Ramsey, M. E., & Gold, J. M. (2007). Overlooking the obvious - A meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Archives of General Psychiatry*, 64(5), 532–542. https://doi.org/10.1001/Archpsyc.64.5.532

Duff, K., Spering, C. C., O'Bryant, S. E., Beglinger, L. J., Moser, D. J., Bayless, J. D., … Scott, J. G. (2011). The RBANS effort index: Base rates in geriatric samples. *Applied Neuropsychology*, 18(1), 11–17. https://doi.org/10.1080/09084282.2010.523354

Duncan, A. (2005). The impact of cognitive and psychiatric impairment of psychotic disorders on the test of memory malingering (TOMM). *Assessment*, 12(2), 123–129. https://doi.org/10.1177/1073191105275512

Egeland, J., Sundet, K., Rund, B. R. R., Asbjrnsen, A., Hugdahl, K., Landr, N. I., … Stordal, K. I. (2003). Sensitivity and specificity of memorydysfunction in schizophrenia: A comparison with major depression. *Journal of Clinical and Experimental Neuropsychology*, *25*(1), 79–93. https://doi.org/10.1076/jcen.25.1.79.13630

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M. E., & Clare, L. (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychology Review*, *15*(2), 73–95. https://doi.org/10.1007/S11065-005-6254-9

Fisher, R. A. (1921). 014: On the "Probable Error" of a coefficient of correlation deduced from a small sample. *Metron, 1,* 3–32.

Foussias, G., Siddiqui, I., Fervaha, G., Mann, S., McDonald, K., Agid, O., … Remington, G. (2015). Motivated to do well: An examination of the relationships between motivation, effort, and cognitive performance in schizophrenia. *Schizophrenia Research*, *166*(1–3), 276–282. https://doi.org/10.1016/j.schres.2015.05.019

Frederick, R. I., Crosby, R. D., & Wynkoop, T. F. (2000). Performance curve classification of invalid responding on the validity Indicator profile. *Archives of Clinical Neuropsychology*, *15*(4), 281–300. https://doi.org/10.1016/S0887-6177(99)00002-5

Fusar-Poli, P., Smieskova, R., Kempton, M., Ho, B., Andreasen, N., & Borgwardt, S. (2013). Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1680–1691. https://doi.org/10.1016/j.neubiorev.2013.06.001

Gierok, S. D., Dickson, A. L., & Cole, J. A. (2005). Performance of forensic and non-forensic adult psychiatric inpatients on the test of memory malingering. *Archives of Clinical Neuropsychology*, *20*(6), 755–760. https://doi.org/10.1016/j.acn.2005.04.008

Glassmire, D. M., Toofanian Ross, P., Kinney, D. I., & Nitch, S. R. (2016). Derivation and cross-validation of cutoff scores for patients with schizophrenia spectrum disorders on WAIS-IV digit span–based performance validity measures. *Assessment*, *23*(3), 292–306. https://doi.org/10.1177/1073191115587551

Gold, J. M., Strauss, G. P., Waltz, J. A., Robinson, B. M., Brown, J. K., & Frank, M. J. (2013). Negative symptoms of schizophrenia are associated with abnormal effort-cost computations. *Biological Psychiatry*, *74*(2), 130–136. https://doi.org/10.1016/j.biopsych.2012.12.022

Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients (research support, non-U.S. Gov't). *Schizophrenia Research*, *78*(2–3), 199–208. https://doi.org/10.1016/j.schres.2005.02.016

Green, M. F., Nuechterlein, K. H., Kern, R. S., Baade, L. E., Fenton, W. S., Gold, J. M., … Marder, S. R. (2008). Functional co-primary measures for clinical trials in schizophrenia: Results from the MATRICS psychometric and standardization study. *American Journal of Psychiatry*, *165*(2), 221–228. https://doi.org/10.1176/Appi.Ajp.2007.07010089

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head injury litigation with the word memory test. *Brain Injury*, *13*(10), 813–819. https://doi.org/10.1080/026990599121205

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the word memory test and medical symptom validity test in groups with severe verbal memory impairment. *Applied Neuropsychology*, *18*(2), 86–94. https://doi.org/10.1080/09084282.2010.523389

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060. https://doi.org/10.1080/02699050110088254

Gur, R. E., Turetsky, B. I., Bilker, W. B., & Gur, R. C. (1999). Reduced gray matter volume in schizophrenia. *Archives of General Psychiatry*, *56*(10), 905–911. https://doi.org/10.1001/archpsyc.56.10.905

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology*, *12*(3), 426–445. https://doi.org/10.1037//0894-4105.12.3.426

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Howe, L. L., Anderson, A. M., Kaufman, D. A., Sachs, B. C., & Loring, D. W. (2007). Characterization of the medical symptom validity test in evaluation of clinically referred memory disorders clinic patients. *Archives of Clinical Neuropsychology*, *22*(6), 753–761. https://doi.org/10.1016/j.acn.2007.06.003

Hunt, S., Root, J. C., & Bascetta, B. L. (2014). Effort testing in schizophrenia and schizoaffective disorder: Validity indicator profile and test of memory malingering performance characteristics. *Archives of Clinical Neuropsychology*, *29*(2), 164–172. https://doi.org/10.1093/arclin/act069

Iverson, G. L., & Tulsky, D. S. (2003). Detecting malingering on the WAIS-III unusual digit span performance patterns in the normal population and in clinical groups. Archives of Clinical Neuropsychology, 18(1), 1–9. https://www.sciencedirect.com/science/article/pii/S0887617701001767?via%3Dihub

Kay, S., Fiszbein, A., & Opler, L. (1987). The positive and negative Syndrom scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261–276. https://doi.org/10.1093/schbul/13.2.261

Keefe, R. S., Eesley, C. E., & Poe, M. P. (2005). Defining a cognitive function decrement in schizophrenia. *Biological Psychiatry*, *57*(6), 688–691. https://doi.org/10.1016/j.biopsych.2005.01.003

Keefe, R. S. E., Poe, M., Walker, T. M., Kang, J. W., & Harvey, P. D. (2006). The schizophrenia cognition rating scale: An interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. *American Journal of Psychiatry*, *163*(3), 426–432. https://doi.org/10.1176/Appi.Ajp.163.3.426

Kirkpatrick, B., Buchanan, R. W., McKenney, P. D., Alphs, L. D., & Carpenter Jr., W. T. (1989). The schedule for the deficit syndrome: An instrument for research in schizophrenia (research support, non-U.S. Gov't research support, U.S. Gov't, P.H.S.). *Psychiatry Research*, *30*(2), 119–123. https://doi.org/10.1016/0165-1781(89)90153-4

Kirkpatrick, B., Messias, E., Harvey, P. D., Fernandez-Egea, E., & Bowie, C. R. (2008). Is schizophrenia a syndrome of accelerated aging? *Schizophrenia Bulletin*, *34*(6), 1024–1032. https://doi.org/10.1093/schbul/sbm140

Kirkpatrick, B., Saoud, J. B., Strauss, G. P., Ahmed, A. O., Tatsumi, K., Opler, M., … Davidson, M. (2018). The brief negative symptom scale (BNSS): Sensitivity to treatment effects. *Schizophrenia Research*, *197*, 269–273. https://doi.org/10.1016/j.schres.2017.11.031

Kirkpatrick, B., Strauss, G. P., Nguyen, L., Fischer, B. A., Daniel, D. G., Cienfuegos, A., & Marder, S. R. (2011). The brief negative symptom scale: Psychometric properties. *Schizophrenia Bulletin*, *37*(2), 300–305. https://doi.org/10.1093/schbul/sbq059

Konstantakopoulos, G., Ploumpidis, D., Oulis, P., Patrikelis, P., Soumani, A., Papadimitriou, G. N., & Politis, A. M. (2011). Apathy, cognitive deficits and functional impairment in schizophrenia. *Schizophrenia Research*, *133*(1–3), 193–198. https://doi.org/10.1093/schbul/sbq059

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Kring, A. M., Gur, R. C., Blanchard, J. J., Horan, W. P., & Reise, S. P. (2013). The clinical assessment interview for negative symptoms (CAINS): Final development and validation. *American Journal of Psychiatry*, *170*(2), 165–172. https://doi.org/10.1176/appi.ajp.2012.12010109

Lange, R. T., Sonal, P., Bhagwat, A., Anderson-Barnes, V., & French, L. M. (2012). Influence of poor effort on neuropsychological test performance in U.S. military personnel following mild traumatic brain injury. *Journal of Clinical and Experimental Neurpsychology*, *34*(5), 453–466. https://doi.org/10.1080/13803395.2011.648175

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology*, *29*(4), 364–373. https://doi.org/10.1093/arclin/acu019

Lezack, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment*. New York, NY: Oxford University Press.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lippa, S. M., Agbayani, K. A., Hawes, S., Jokic, E., & Caroselli, J. S. (2015). Effort in acute traumatic brain injury: Considering more than pass/fail. *Rehabilitation Psychology*, *59*(3), 306–312. https://doi.org/10.1037/a0037217

Lippa, S. M., Lange, R. T., French, L. M., & Iverson, G. L. (2018). Performance validity, neurocognitive disorder, and post-concussion symptom reporting in service members with a history of mild traumatic brain injury. *Archives of Clinical Neuropsychology*, *33*(5), 606–618. https://doi.org/10.1093/arclin/acx098

Llerena, K., Strauss, G. P., & Cohen, A. S. (2012). Looking at the other side of the coin: A meta-analysis of self-reported emotional arousal in people with schizophrenia. *Schizophrenia Research*, *142*(1–3), 65–70. https://doi.org/10.1016/j.schres.2012.09.005

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the test of memory malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, *34*(1), 88–119. https://doi.org/10.1080/13854046.2019.1637027

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, *34*(1), 114–131. https://doi.org/10.1093/arclin/acy012

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, *29*(3), 308–318. https://doi.org/10.1080/13803390600693607

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology*, *27*(1), 10–22. https://doi.org/10.1093/arclin/acr084

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exeggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1094–1102. https://doi.org/10.1076/jcen.24.8.1094.8379

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moore, R. C., Davine, T., Harmell, A. L., Cardenas, V., Palmer, B. W., & Mausbach, B. T. (2013). Using the repeatable battery for the assessment of neuropsychological status (RBANS) effort index to predict treatment group attendance in patients with schizophrenia. *Journal of the International Neuropsychological Society*, *19*(2), 198–205. https://doi.org/10.1017/s1355617712001221

Morra, L. F., Gold, J. M., Sullivan, S. K., & Strauss, G. P. (2015). Predictors of neuropsychological effort test performance in schizophrenia. *Schizophrenia R esearch*, *162*(1–3), 205–210. https://doi.org/10.1016/j.schres.2014.12.033

National Heart, Lung, & Blood Institute. (2014). Quality assessment tool for observational cohort and cross-sectional studies. Bethesda: National Institutes of Health, Department of Health and Human Services. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Nelson, N. W., Hoelze, J. B., McGuire, K. A., Ferrier-Auerbach, A. G., Charlesworth, M. J., & Sponheim, S. R. (2010). Evaluation context impacts neuropsychological performance of OEF/OIF veterans with reported combat-related concussion. *Archives of Clinical Neuropsychology*, *25*(8), 713–723. https://doi.org/10.1093/arclin/acq075

Overall, J. E., & Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychological Reports*, *10*(3), 799–812. https://doi.org/10.2466/pr0.1962.10.3.799

Paradis, C. M., Solomon, L. Z., Owen, E., & Brooker, M. (2013). Detection of cognitive malingering or suboptimal effort in defendants undergoing competency to stand trial evaluations. *Journal of Forensic Psychology Practice*, *13*(3), 245–265. https://doi.org/10.1080/15228932.2013.803374

Pivovarova, E., Rosenfeld, B., Dole, T., Green, D., & Zapf, P. (2009). Are measures of cognitive effort and motivation useful in differentiating feigned from genuine psychiatric symptoms? *International Journal of Forensic Mental Health*, *8*(4), 271–278. https://doi.org/10.1080/14999011003635514

Raffard, S., Capdevielle, D., Boulenger, J.-P., Gely-Nargeot, M.-C., & Bayard, S. (2014). Can individuals with schizophrenia be instructed to deliberately feign memory deficits? *Cognitive Neuropsychiatry*, *19*(5), 414–426. https://doi.org/10.1080/13546805.2014.896251

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the test of memory malingering (TOMM). *Psychological Assessment*, *10*(1), 10–20. https://doi.org/10.1037/1040-3590.10.1.10

Reichenberg, A., & Harvey, P. D. (2007). Neuropsychological impairments in schizophrenia: Integration of performance-based and brain imaging findings. *Psychological Bulletin*, *133*(5), 833–858. https://doi.org/10.1037/0033-2909.133.5.833

Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2018). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology*, *25*(3), 219–230. https://doi.org/10.1080/23279095.2017.1278602

Rohling, M. L., Binder, L. M., & Langhinrichsen-Rohling, J. (1995). Money matters: A meta- analytic review of the association between financial compensation and the experience and treatment of chronic pain. *Health Psychology*, *14*(6), 537–547. https://doi.org/10.1037/0278-6133.14.6.537

Rosenberg, M. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, *59*(2), 464–468. https://doi.org/10.1111/j.0014-3820.2005.tb01004.x

Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist*, *25*(3), 437–453. https://doi.org/10.1080/13854046.2011.556668

Sieck, B. C., Smith, M. M., Duff, K., Paulsen, J. S., & Beglinger, L. J. (2013). Symptom validity test performance in the Huntington disease clinic. *Archives of Clinical Neuropsychology*, *28*(2), 135–143. https://doi.org/10.1093/arclin/acs109

Singhal, A., Green, P., Ashaye, K., Shankar, K., & Gill, D. (2009). High specificity of the medical symptom validity test in patients with very severe memory impairment. *Archives of Clinical Neuropsychology*, *24*(8), 721–728. https://doi.org/10.1093/arclin/acp074

Slick, D., Hopp, G., Strauss, E., & Thompson, G. (1997). *The Victoria symptom validity test*. Odessa: Psychological Assessment Resources. https://doi.org/10.1080/01688639608408313

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561. https://doi.org/10.1076/1385-4046(199911)13:04; 1-y;ft545

Smith, K., Boone, K., Victor, T., Miora, D., Cottingham, M., Ziegler, E., … Wright, M. (2014). Comparison of credible patients of very low intelligence and non-credible patients on neurocognitive performance validity indicators. *The Clinical Neuropsychologist*, *28*(6), 1048–1070. https://doi.org/10.1080/13854046.2014.931465

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Eepidemiology*, *53*(11), 1119–1129. https://doi.org/10.1016/s0895-4356(00)00242-0

Stevens, A., Schneider, K., Liske, B., Hermle, L., Huber, H., & Hetzel, G. (2014). Is subnormal cognitive performance in schizophrenia due to lack of effort or to cognitive impairment? German Journal of Psychiatry, 17(1), 1–9. https://www.semanticscholar.org/paper/Is-Subnormal-Cognitive-Performance-in-Schizophrenia-Stevens-Schneider/0d4c8f0ab3c3dc88aa52d7e8718f9fa4e29d8d68

Strauss, G. P., & Chapman, H. C. (2018). Preliminary psychometric properties of the brief negative symptom scale in youth at clinical high-risk for psychosis. *Schizophrenia Research*, *193*, 435–437. https://doi.org/10.1016/j.schres.2017.07.051

Strauss, G. P., & Gold, J. M. (2016). A psychometric comparison of the clinical assessment interview for negative symptoms and the brief negative symptom scale. *Schizophrenia Bulletin*, *42*(6), 1384–1394. https://doi.org/10.1093/schbul/sbw046

Strauss, G. P., Esfahlani, F. Z., Galderisi, S., Mucci, A., Rossi, A., Bucci, P., … Sayama, H. (2018a). Network analysis reveals the latent structure of negative symptoms in schizophrenia. *Schizophrenia Bulletin*, *45*(5), 1033–1041. https://doi.org/10.1093/schbul/sby133

Strauss, G. P., Hong, L. E., Gold, J. M., Buchanan, R. W., McMahon, R. P., Keller, W. R., … Kirkpatrick, B. (2012a). Factor structure of the brief negative symptom scale. *Schizophrenia Research*, *142*(1–3), 96–98. https://doi.org/10.1016/j.schres.2012.09.007

Strauss, G. P., Keller, W. R., Buchanan, R. W., Gold, J. M., Fischer, B. A., McMahon, R. P., … Kirkpatrick, B. (2012b). Next-generation negative symptom assessment for clinical trials: Validation of the brief negative symptom scale. *Schizophrenia Research*, *142*(1–3), 88–92. https://doi.org/10.1016/j.schres.2012.10.012

Strauss, G. P., Morra, L. F., Sullivan, S. K., & Gold, J. M. (2015). The role of low cognitive effort and negative symptoms in neuropsychological impairment in schizophrenia. *Neuropsychology*, *29*(2), 282–291. https://doi.org/10.1037/neu0000113

Strauss, G. P., Nuñez, A., Ahmed, A. O., Barchard, K. A., Granholm, E., Kirkpatrick, B., … Allen, D. N. (2018b). The latent structure of negative symptoms in schizophrenia. *JAMA Psychiatry*, *75*(12), 1271–1279. https://doi.org/10.1001/jamapsychiatry.2018.2475

Strauss, G. P., Vertinski, M., Vogel, S. J., Ringdahl, E. N., & Allen, D. N. (2016). Negative symptoms in bipolar disorder and schizophrenia: A psychometric evaluation of the brief negative symptom scale across diagnostic categories. *Schizophrenia Research*, *170*(2–3), 285–289. https://doi.org/10.1016/j.schres.2015.12.014

Sugawara, N., Maruo, K., Sugai, T., Suzuki, Y., Ozeki, Y., Shimoda, K., … Yasui-Furukori, N. (2018). Prevalence of underweight in patients with schizophrenia: A meta-analysis. *Schizophrenia Research*, *195*, 67–73. https://doi.org/10.1016/j.schres.2017.10.017

Sullivan, E. V., Shear, P. K., Lim, K. O., Zipursky, R. B., & Pfefferbaum, A. (1996). Cognitive and motor impairments are related to gray matter volume deficits in schizophrenia. *Biological Psychiatry*, *39*(4), 234–240. https://doi.org/10.1016/0006-3223(95)00135-2

Teichner, G., & Wagner, M. T. (2004). The test of memory malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, *19*(3), 455–464. https://doi.org/10.1016/S0887-6177(03)00078-7

Tombaugh, T. N. (1996). *Test of memory malingering (TOMM)*. New York: Multi-Health Systems, Inc..

Tombaugh, T. N. (1997). The test of memory malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, *9*(3), 260–268. https://doi.org/10.1037/1040-3590.9.3.260

Velligan, D. I., Kern, R. S., & Gold, J. M. (2006). Cognitive rehabilitation for schizophrenia and the putative role of motivation and expectancies. *Schizophrenia Bulletin*, *32*(3), 474–485. https://doi.org/10.1093/schbul/sbj071

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03

Victor, T. L., & Boone, K. B. (2007). Identification of feigned mental retardation. *Assessment of feigned cognitive impairment: A neuropsychological perspective*, 310–345. https://doi.org/10.1076/jcen.25.7.979.16481

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLoS One*, *8*(7), e68210. https://doi.org/10.1371/journal.pone.0068210

Whearty, K. M., Allen, D. N., Lee, B. G., & Strauss, G. P. (2015). The evaluation of insufficient cognitive effort in schizophrenia in light of low IQ scores. *Journal of Psychiatry Research*, *68*, 397–404. https://doi.org/10.1016/j.jpsychires.2015.04.018

Whitney, K. A., Shepard, P. H., Mariner, J., Mossbarger, B., & Herman, S. M. (2010). Validity of the Wechsler test of adult Reading (WTAR): Effort considered in a clinical sample of U.S. military veterans. *Applied Neuropsychology*, *17*(3), 196–204. https://doi.org/10.1080/09084282.2010.499787

Whiteney, K. A., Shepard, P. H., Williams, A. L., Davis, J. J., & Adams, K. M. (2009). The medical symptoms validity test in the evaluation of operation Iraqi freedom/operation enduring freedom soldiers: A preliminary study. *Archives of Clinical Neuropsychology*, *24*(2), 145–152. https://doi.org/10.1093/arclin/acp020

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A meta-analysis of cognitive remediation for schizophrenia: Methodology and effect sizes. *American Journal of Psychiatry*, *168*(5), 472–485. https://doi.org/10.1176/appi.ajp.2010.10060855

Zipursky, R. B., Lambe, E. K., Kapur, S., & Mikulis, D. J. (1998). Cerebral gray matter volume deficits in first episode psychosis. *Archives of General Psychiatry*, *55*(6), 540–546. https://doi.org/10.1001/archpsyc.55.6.540