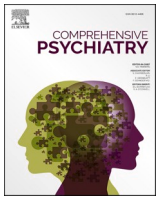
Comprehensive Psychiatry 129 (2024) 152444

Contents lists available at ScienceDirect

Comprehensive Psychiatry

journal homepage: www.elsevier.com/locate/comppsych



Proverb comprehension in schizophrenia: A comprehensive review and meta-analysis

Anne Felsenheimer [a,b], Alexander M. Rapp [a,*]

[a] Department of Psychiatry and Psychotherapy, University of Tuebingen, Calwerstrasse 14, 72076 Tuebingen, Germany
[b] Max Planck School of Cognition, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany

ARTICLE INFO

Keywords:
Metaphor
Thought disorder
Language
Psychosis
Social cognition
Depression

ABSTRACT

Background: Examination of proverb comprehension has a long tradition in clinical diagnostics of individuals with schizophrenia (iSCZ). Deficits in the comprehension are considered common. Interpretations of proverbs are traditionally measured by their degree of abstraction and concreteness ('literalness'), but iSCZ's responses may also be illogical or 'bizarre'. Experimental research on proverb comprehension starts in the 1940s. Since then, the specificity of proverb tests has often been questioned, but has never been the subject of a meta-analysis. The aim of this meta-analysis is to include all experimental research, including historical studies, that meets quality criteria and compares the responses to proverbs in iSCZ with those in healthy controls (HC) or clinical controls (CC).

Methods: PubMed, Web of Science, and PsycInfo databases were searched. After coding 121 articles, 27 (median publication year 1982) were included and multi-level meta-analyses performed. Moderator analyses were performed on response format (multiple-choice vs. verbal responses), proverb test, scoring method, language, acute vs. chronic stage of iSCZ, time of publication, clinical vs. healthy control group, age, IQ/education, and gender. Publication bias was assessed using funnel plots, trim and fill method and Egger's test.

Results: The search identified 27 eligible studies for inclusion. Studies were published between 1956 and 2020 and predominantly older than 30 years (median: 1982). The Gorham Proverbs Test was the most established test and predominantly conducted in English. CC mostly consisted of depressive disorders. Pooled estimates yielded statistically significant less abstract ($g = -1.00$; 95%CI, −1.34 to −1.67), more concrete ($g = 0.69$; 95%CI, 0.35–1.03), and more bizarre ($g = 1.08$; 95%CI, 0.74–1.41) responses in iSCZ compared to controls. The type of control group moderated all three effects, with greater differences of iSCZ compared to HC than to CC in abstraction and bizarreness, and no significant group difference between iSCZ and CC in concreteness. Meta-regressions indicated IQ/education and age as possible sources of variability in abstraction and bizarreness.

Conclusions: While lower abstraction and higher bizarreness seems a characteristic of iSCZ, the diagnostic specificity of a concrete response was astonishingly low. The lack of a unified definition for concretism and limited consideration of cultural diversity contributed to these complex findings. Future research should focus on exploring the qualitative aspects of proverb comprehension and the association between symptomatology types and misinterpretations to improve diagnostic accuracy.

1. Introduction

Examination of figurative language in individuals with schizophrenia (SCZ) has a long tradition among clinicians and researchers [1–4]. One aspect of interest is concretism, which refers to the tendency to interpret nonliteral or "figurative" expressions such as proverbs, idioms or metaphors in a literal manner. This phenomenon was described by Bleuler in his initial characterization of the disorder [5], and subsequent studies have consistently associated it with SCZ, indicating its potential as a valuable clinical marker [5–8]. Early attempts to provide

Abbreviations: CC, clinical control group; CI, confidence interval; GPT, Gorham Proverbs Test; HC, healthy control group; iSCZ, individuals with schizophrenia; M, mean; MC, multiple-choice; PANSS, Positive and Negative Syndrome Scale; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analysis; RC, reference category; SCZ, schizophrenia; SD, standard deviation; VE, verbal explanation.

* Corresponding author at: Department of Psychiatry and Psychotherapy, University of Tuebingen, Calwerstraße 14, Tuebingen, Germany.
E-mail address: alexander.rapp@med.uni-tuebingen.de (A.M. Rapp).

https://doi.org/10.1016/j.comppsych.2023.152444

Available online 17 December 2023

experimental explanations for concretism in SCZ were made by Goldstein [9] who proposed that individuals with SCZ struggle to form abstract concepts, leading them to adhere to the literal or "concrete" meanings of expressions. Today, this assumed deficit in abstraction is still the concept behind vocabulary tests in widely used scales, such as the Positive and Negative Syndrome Scale (PANSS) [10]. In that way, proverbs are used to evaluate abstract reasoning and thought disorder [1,11,12]. Building upon Goldstein's hypothesis, Benjamin [13] advocated for a systematic assessment of proverb comprehension in individuals with SCZ. In response to this call, the Gorham Proverbs Test (GPT) [2,14] was introduced as the first standardized test, employing multiple-choice or free-response formats to evaluate concretism in this population. Although other tests have been developed [6,15–17], the English language GPT represents the most prevalent paradigm for empirically assessing concretism. Since the introduction of the GPT, >10 different scoring systems for verbal responses have been developed [1,2,13,18–26]. The original objective of the GPT aligns with the idea that a deficiency in abstraction ability contributes to literal thinking [9]. Consequently, the GPT aims to capture two key aspects: the level of abstraction and the level of concreteness in a patient's responses. A successful interpretation of a proverb like "the sun shines on us all alike," includes an abstraction of the literal words to detect the symbolic meaning (e.g., "All are created equal."). In a concrete response the proverb is understood as a factual statement, i.e., the patient sticks to the literal words ["The sun shines on everybody"; 27]. Importantly, the symbolic meaning still needs to be logically connected with the literal meaning [27,28]. If this connection is lost, it can result in a response often labelled as bizarre, illogical, or idiosyncratic [18,19,22,23]. For example, according to Marengo, Harrow [22], a patient's bizarre response to "Rome wasn't built in a day" would be "It's love, I think of it as love. I have to work towards love, and love must work towards me. This must be performed gradually.". In sum, the cognitive process of understanding proverbs necessitates striking a delicate balance between widening associations while maintaining a logical connection to the literal statement [29].

But its comprehension can be influenced by several other factors. For example, proverb comprehension may be influenced by education or intelligence [30–32], age [30,33,34], or the chronic or acute stage of a patient [30,35–37]. However, the evidence regarding these factors is inconclusive. Some studies have shown that patients' proverb interpretation is independent of intelligence [38,39], with the pragmatic competence of individuals with SCZ remaining even when controlling for intelligence [40,41]. Other studies have shown that proverb comprehension in individuals with SCZ is at least partly explained by intelligence [19,31,42]. In addition to participant characteristics, task demands may also influence performance. Bambini et al. [43] showed that verbal responses to proverbs were more difficult for individuals with SCZ than multiple-choice formats. Different response formats require distinct brain correlates [44,45] with a significantly greater contribution of the right hemisphere to the multiple-choice format [46,47].

In line with these possible confounding factors, the effect size and reliability of proverb tests have been debated early on [1,48,49], most prominently and compellingly by Nancy Andreasen [1], emphasizing the lack of differential diagnostic specificity compared to clinical control groups, such as depression [50,51]. In particular, it remains unclear which aspect, i.e., a lack of abstraction, a tendency to interpret concretely, or a bizarre response, may be more specific to individuals with SCZ. Thus, it surprising that, despite several narrative reviews (e.g. [1,3,52,53]), no meta-analysis has been conducted to evaluate these doubts.

Given the long tradition and remaining significance of proverb tests in daily clinical assessments in identifying thought disorder, conducting a meta-analysis is crucial. By systematically examining the existing research, the current meta-analysis aims to provide a rigorous evaluation of the effect size and diagnostic specificity of proverb tests in individuals with schizophrenia. This meta-analysis aims to address the existing gap in the schizophrenia literature by providing a comprehensive understanding of the type of response (abstract, concrete, and bizarre) observed in proverb tests among individuals with schizophrenia. Furthermore, it seeks to examine the role of different control groups (clinical control group vs. healthy control group) in assessing the diagnostic specificity of proverb tests in this population.

2. Methods

2.1. Literature search

We performed a systematic literature search of experimental proverb comprehension studies on individuals with SCZ based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [54] (see Fig. 1). We searched PubMed, Web of Science, and PsycINFO databases for literature published between 1921 and 2021. Search terms were "proverb", "proverbs" and "proverb interpretation" in combination with "schizophrenia", "schizophrenics", "psychosis" and "concretism". The reference lists of these articles were screened to identify relevant studies. In addition, Web of Science citation lists of seminal proverb tests (e.g.([55,56]) were screened. The selection procedures were independently performed by two researchers (A.F. and A. R.). Disagreements were resolved by consensus after discussion.

Studies fulfilling the following criteria were included in this review and meta-analysis:

1. Empirical studies specifically focusing on individuals diagnosed with schizophrenia.
2. Clinical or healthy control group included in the study.
3. Participants limited to adults (i.e., above 18 years old).
4. Reporting of at least one of the outcomes of abstraction, concreteness, and bizarreness of the response.
5. Utilization of a minimum of three verbally or written presented proverbs as stimuli.
6. Availability of sufficient outcome data to allow for the calculation of effect sizes.
7. Studies had to be published in English, French, Spanish, or German because at least one of the authors mastered one or more of these languages.

Some systems propose additional scales, such as more refined subscales to assess abstraction and concreteness, e.g., distinguishing between "abstract correct" and "abstract false" responses [57,58,59]. In these cases, the scales chosen were those that were most consistent with the definitions and examples of abstraction, concreteness, and bizarreness provided in the introduction. Likewise, differing classification as acute or chronic stage of the disorder, e.g., Kantor, Wallner [60] distinction of process vs. reactive used in Watson [37] or the paranoid vs. non-paranoid distinction [38], were fitted to these categories.

Case studies, studies on other nonliteral expressions (e.g., idioms, metaphors, hyperbole, irony or sarcasm), and studies measuring other performance metrics (e.g., response time or word count) were excluded. After removing duplicate articles and coding 121 full articles, 27 studies provided sufficient data for inclusion in the meta-analysis.

2.2. Sample and study characteristics

The characteristics of each study and sample are presented in Table 1. Studies were published between 1956 and 2020 and were predominantly older than 30 years (median: 1982). Notably, CC consisted mostly of individuals with depressive disorders. The GPT [2,14] was the oldest and most established test. The proverbs used as stimuli covered a wide range of item numbers. They included stimuli from the GPT [2,57,30,31,37,50,51,58,61–64] and researchers' own selections [26,42,65], subsets of other test batteries [16,66] and compilations by

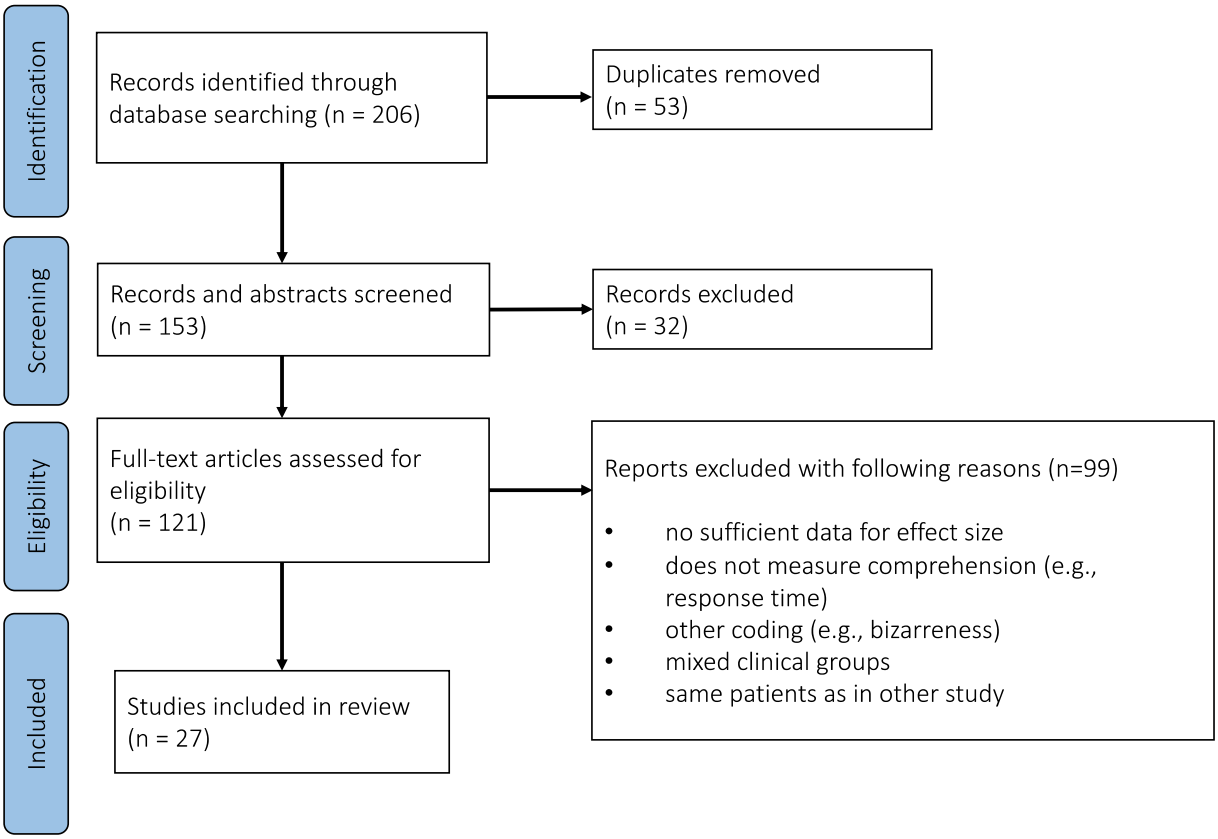


Fig. 1. Literature search based on Preferred Reporting Items for Systematic Reviews and Meta Analysis (PRISMA) guideline.

other scholars [6,13]. Different sets of proverbs were often combined with different scoring systems for verbal responses, which mostly varied in dimensionality (categorical vs. dimensional), constructs (abstraction vs. concreteness vs. bizarre-idiosyncratic) and semantic parts of the proverb. Some variation in the scoring systems employed reflected different conceptualizations of abstraction and concreteness. For example, Gorham's own manual [67] initially scored concreteness and abstraction separately, later viewing them as endpoints on a unified abstract reasoning continuum [14]. In line with this view, some authors combined both concrete and abstract responses into a single score, with concrete responses indicating lower abstraction [18,66,68]. In other cases, scoring manuals allowed for independent evaluation of accuracy and abstraction so that responses could be scored as abstract but not necessarily correct [15,57,69]. In addition, some authors divided the proverb into two logical parts ("Rome" and "built in a day"; [20]), which were scored separately. Notably, Shimkunas [70] introduced the first scoring system for "autistic" responses, characterized by their peculiar, idiosyncratic, bizarre, inappropriate, or tangentially relevant nature to the proverb's meaning. Subsequent systems adopted the term "bizarre-idiosyncratic" [22,23] to describe similar responses. Detailed descriptions of the scoring systems utilized in the meta-analysis are available in the Supplementary Table 1.

2.3. Data analysis

Analyses were performed using the *metafor* package [71] in R (version 4.3.1). The procedure followed the analysis suggested by Assink and Wibbelink [72]. Means and standard deviations were extracted from the studies or calculated based on the reported data. The standardized mean difference was calculated using the *escalc* function and corrected for small samples by transformation to Hedges' g [73].

Most studies reported more than one effect size (e.g., using two tasks or two groups of patients). Same study effect sizes are typically dependent as they are derived from equivalent participants, instruments, or conditions within the study, which may artificially reduce heterogeneity and increase the number of false positives. For this reason, we used a three-level meta-analysis instead of the originally planned regular meta-analysis, which does not model within-study variance. The three levels are used to account for the nested nature of the data: individuals in the primary study (level 1) and same-study effect sizes that are nested within the study from which they originate (level 2), and that are aggregated to form the overall "true" effect size between studies (level 3). This permits effects to differ between participants (level 1), outcomes (level 2), and studies (level 3). This allows all available information to be used. In particular, it allows the variance of effect sizes to be estimated not only between studies (level 3), but also within studies (level 2), as an indicator of the heterogeneity introduced at each level.

For each response score (abstraction, concreteness and bizarreness), we first modelled a three-level meta-analysis with random effects for the individual effect size (level 2) and study (level 3) each. To determine the significance of between-study and within-study variance, we performed two independent one-tailed log-likelihood ratio tests using the *anova* function. In these tests, we compared the deviance of the full model to that of the model with either within-study variance or between-study variance manually fixed to zero. Consequently, significance indicated a decrease in model fit when within-study or between-study variance was excluded.

For all analyses, the significance level was set at 5%. The Knapp-Hartung-correction [74] was applied, which uses a t-distribution to test the coefficients of the regression model. Once the predictors for the moderator analyses were added to the model, this correction applies an F-distribution as an omnibus test of all included coefficients.

For the moderator analyses, the continuous variables were centered around the mean and the categorical variables were dummy-coded. In the moderator models, categories are tested against a reference category

Table 1
Main characteristics of included studies.

Study	Patient groups				Control groups					Study characteristics				
	Diagnosis	n	f/m	Age years (mean)	Type	n	f/m	Age years (mean)	Matched variables	Proverbs from	Number of proverbs	Scores assessed	Response format	Language
Ahmad (1986)	SCZ	20	0/20	/	HC	20	0/20	/	age, gender[a], education, socio-economic status, religion	GPT	12	A, C	VE	Hindi
					CC (depression)	20	0/20	/						
Bambini et al. (2020)	SCZ	47	18/29	43.7	HC	39	22/17	39.7	age, education, gender	Arcara & Bambini (2016)		A	VE, MC	Italian
Braff et al. (1974)	SCZ	24	14/10	31.2	HC	24	13/11	26.3	age, gender, intelligence	GPT	12	B, C	VE	English
					CC (depression)	19	14/5	45.8						
Braff et al. (1983)	SCZ	20	/	27.2	HC	19	/	28.4	/	GPT	12	B, C	VE	English
					CC (depression)	15	/	37.3						
Braff et al. (1988)	SCZ	19	/	27.5	HC	16	/	40.4	/	GPT	12	A, B	VE	English
					CC (depression)	14	/	28.9						
Brattemo (1962)	SCZ chronic	30	/	39.8	HC	30	/	39.1	age, intelligence	GPT	12 (VE) 40 (MC)	A, C	VE (only CC), MC	Swedish
	SCZ acute	30	/	37.8	CC (depression)	30	/	40.2						
Brattemo (1967)	SCZ chronic	33	/	39.9	HC (33)	33	/	39.6	age, intelligence	GPT	12 (VE) 40 (MC)	A	VE, MC	Swedish
	SCZ acute	33	/	40.1	CC (depression)	33	/	42.2						
					CC (head injury)	33	/	41.3						
Bömmer and Brüne (2006)	"delusional disorder"	21	14/7	51.6	HC	22	12/10	45.7	sex, intelligence, education	Barth & Küfferle (2001)	14	A, C	MC	German
Brüne and Bodenstein (2005)	SCZ	31	8/23	38.6	HC	21	11/10	33.9	age, gender	Barth & Küfferle (2001)	14	A	MC	German
Carter (1986)	SCZ	25	17/8	33.4	HC	25	20/5	35.2	age, sex, education	GPT	12	A, B, C	VE	English
					CC (depression)	25	20/5	38.8						
Chapin et al. (1996)	SCZ	40	10/30	27.7	HC	40	10/30	27.3	age, sex, race	GPT	40	A, C	MC	English
					CC (bipolar, depression)	20	5/15	31.9						
Glicksohn et al. (2001)	SCZ	8	4/4	27.5	HC	10	5/5	28	age, education	GPT/ Benjamin (1944)	15	A	MC	Hebrew
Gorham (1956)	SCZ	100	0/100	/	HC	100	0/100	/	sex, intelligence, education	GPT	12 (VE) 40 (MC)	A, C	VE, MC	English
Haas et al. (2015)	SCZ	15	/	23.5	HC	7	/	28	age	Barth & Küfferle (2001)	14	A	VE	English
Hamlin et al. (1965)	closed ward SCZ	14	0/14	39.1	CC (non-SCZ patients)	14	0/14	41	age, education	Blaufarb (1962)	57	A	VE	English
	open ward SCZ	14	0/14	38.6										
Hamlin and Folsom (1977)	paranoid SCZ	12	0/12	42.8	HC	12	0/12	41.75	age, education	Blaufarb (1962)	57	A	VE	English
	non-paranoid SCZ	12	0/12	37.9	CC (neurotic)	12	0/12	46.2						

(*continued on next page*)

Table 1 (*continued*)

Study	Patient groups				Control groups					Study characteristics				
	Diagnosis	n	f/m	Age years (mean)	Type	n	f/m	Age years (mean)	Matched variables	Proverbs from	Number of proverbs	Scores assessed	Response format	Language
					CC (lesion)	12	0/12	44.6						
Harrison et al. (1972)	acute thought disordered SCZ	12	/	/	CC (bipolar)	12	/	/	education	self-selected	6	A	VE	English
	acute non-thought disordered SCZ	12	/	/										
	chronic thought disordered SCZ	12	/	/										
	chronic non-thought disordered SCZ	12	/	/										
Harrow et al. (1972)	SCZ	45	23/22	23	CC (depression, personality disorder, other)	47	14/33	23	gender, intelligence	Benjamin (1944)	14	A, B, C	VE	English
Kiang et al. (2007)	SCZ	18	5/13	46.3	HC	18	6/12	43.3	age, sex, parental socioeconomic status	Delis et al. (2001)	8	A	VE	English
Lewis et al. (1959)	SCZ	30	17/13	31.7	HC	30	16/14	33.6	age, sex, education	self-selected	16	A	MC	English
Mossaheb et al. (2014)	SCZ, schizoaffective	40	25/15	39.6	HC	40	24/16	42.9	age, sex, education	self-selected	12	A	VE, MC	German
Phillips et al. (1980)	SCZ	23	/	27	CC (non-SCZ patients)	36	/	27	/	GPT	12	A, C	VE, MC	English
Reed (1968)	SCZ	18	/	/	HC	18	/	/	social class, education	self-selected	20	A	VE	English
Reich (1981)	SCZ	21	8/13	31	HC	22	12/10	32.5	age, sex, race, education level	Benjamin (1944)	4	A	VE	English
					CC (bipolar)	22	8/14	35.4						
Sponheim et al. (2003)	SCZ	23	0/23	41.8	HC	29	0/29	37.8	gender, parental socioeconomic status	GPT	12	A, C	VE	English
Watson (1973)	*reactive[b] (acute) SCZ matched	40	0/40	39.1	matched HC	20	0/20	42.8	age, education, intelligence	GPT	12	A, B	VE	English
	reactive (acute) SCZ not matched	48	0/48											
	*process (chronic) SCZ matched	40	0/40	38.1	not matched HC	24	0/24							
	process (chronic) SCZ not matched	48	0/48											

Note: SCZ = schizophrenia, / = not reported, A = abstract, C = concrete, B = bizarre, VE = verbal explanation, MC = multiple-choice, HC = healthy control group, CC = clinical control group, GPT = Gorham Proverbs Test.

* Selected sample for analyses.

[a] Terms sex and gender are listed as used in the quoted study.

[b] According to the concept of process and reactive schizophrenia by Kantor (1953).

(RC). In addition, we tested each coefficient of the moderators against zero. First, we included differences in mean age (years), intelligence/education and percentage of females as continuous predictors in the model. Additionally, we conducted categorical moderator analyses for type of controls (HC vs. CC), language, response format (verbal explanation vs. multiple choice) and stage of individuals with SCZ (acute vs. chronic vs. mixed). Only four of the 27 included studies reported the participants' race/ethnicity. Two studies had samples of Caucasian [64] or "British" cultural background [65] and two reported the percentage of people of color [26,63], without further discrimination. Thus, this category was not further analyzed in the meta-analysis. Apart from the planned moderator analyses, we included factors, that came up as relevant during the coding of studies: Considering the possible inferiority of the diagnostic criteria prior to the introduction of DSM-III in terms of reliability and precision of differential diagnosis [75], we incorporated the timing of the study—either before or after the implementation of DSM-III in 1980. Moreover, we added the source of the proverbs that were selected for testing, as well as the scoring system that was used to identify responses as abstract, concrete, or bizarre. These analyses should be regarded as exploratory.

Publication bias was investigated using funnel plots, Egger's regression analysis [76] and the trim and fill method using the *trimfill* function, the latter not being part of the original analysis plan but added for robustness. For this, we aggregated the individual effect sizes on the study level with the using the *aggregate* function. In case of an asymmetrical funnel plot, the trim and fill method imputes effect size estimates from studies that are missing for symmetry [77].

3. Results

3.1. Abstraction

The meta-analysis on the difference between the abstraction of individuals with SCZ and control participants in interpreting proverbs was based on 55 effect sizes derived from 24 studies (see Table 2 and Fig. 2). It yielded a significant overall effect (Hedges' $g = -1.004$ [95% CI = −1.34 to −0.67]; $p < .001$), indicating that individuals with SCZ provided fewer abstract interpretations of proverbs than participants in control groups. The effect can be regarded as large [78]. The log-likelihood ratio tests revealed significant variance within studies (level 2: χ^2 (2) = 47.178, $p < .001$) and between studies (level 3: χ^2 (2) = 8.246, $p = .001$), indicating significant heterogeneity of effect sizes. Thus, we continued with the moderator analysis.

The results of the moderator analysis are presented in Table 3. The meta-regression indicated significant effects of IQ/education and age on the abstract interpretation of proverbs. Older age was related to fewer abstract responses. We found that the difference in abstract responses was smaller with a smaller difference in IQ/education between individuals with SCZ and controls. However, deficits remained even in samples matched for IQ/education (see Supplementary Fig. 4 for a scatter plot). Moreover, when individuals with SCZ were older than controls, the difference in abstraction was greater than in studies with younger individuals with SCZ than controls (Supplementary Fig. 4). The results revealed significant moderating effects of the type of control group (CC vs. HC) and language. Despite both groups showed significant differences to individuals with SCZ, these effects were smaller when the control group consisted of clinical individuals. Additionally, studies conducted in Hindi showed significantly lower abstraction scores for in individuals with SCZ when compared to the reference category (English). It is important to note that the majority of studies were conducted in English, which captures most of the variability, while only one study was conducted in Hindi [61].

3.2. Concreteness

The difference between the concreteness of proverb interpretation in individuals with SCZ and control participants was based on 21 effect sizes derived from 11 studies (see Table 2 and Fig. 3). It revealed an overall significant medium effect (Hedges' $g = 0.690$ [95% CI = 0.35–1.03]; $p < .001$), indicating that individuals with SCZ provided more abstract interpretations of proverbs than participants in control groups. While log-likelihood ratio tests revealed significant within-study variance (level 2: χ^2 (2) = 31.860, $p < .001$), this heterogeneity did not extend between studies (level 3: χ^2 (2) = 0.034, $p = .427$). However, opting for the reduced model in our moderator analyses would imply an assumption of no random variability at the study level. Allowing for random variability between studies is more consistent with the nature of meta-analysis and ensures a higher level of robustness in our findings by accounting for potential between-study variance. Hence, despite its conservative nature, we chose to use the full model for our moderator analyses.

The results of the moderator analysis are presented in Table 4. Age, IQ/education, and percentage of females showed no significant moderating effects on the concreteness of responses. The type of control group significantly moderated the overall effect, with significant differences between individuals with SCZ and HC, but not CC. There was a marginally statistically significant effect of language. Again, the study conducted in Hindi tended to show greater differences between groups when compared to studies conducted in English.

3.3. Bizarreness

The difference between the bizarreness of proverb interpretation in individuals with SCZ and control participants was based on 11 effect sizes derived from 7 studies (see Table 2 and Fig. 4). It yielded an overall significant large effect (Hedges' $g = 1.078$ [95% CI = 0.74–1.42]; $p < .001$), indicating that individuals with SCZ gave more bizarre interpretations of proverbs than participants in control groups. While log-likelihood ratio tests revealed significant within-study variance (level 2: χ^2 (2) = 4.366, $p < .019$), this heterogeneity did not extend between studies (level 3: χ^2 (2) = 0.00, $p = 1.000$). Adopting the same considerations as for the model for concreteness, we continued with the three-level model for the moderator analyses.

The moderator analyses are presented in Table 5. There was a significant moderating effect of age, IQ/education and type of control group. The difference in bizarreness tended to be smaller in studies where the education/IQ of individuals with SCZ was higher than that of controls (Supplementary Fig. 7). In addition, studies in which individuals with SCZ were older than controls tended to show a greater difference in bizarreness than studies in which they were younger (Supplementary Fig. 8). Although both groups showed significant differences from individuals with SCZ, these effects were smaller when the

Table 2
Overall differences in proverb interpretation between individuals with schizophrenia and control groups.

Outcome	*s*	*k*	Mean g	95% CI	*p*	σ^2 level2	σ^2 level3	% Var. level 1	% Var. level 2	% Var. level 3
Abstraction	24	55	−1.004	−1.337 – −0.671	<0.001***	0.265***	0.431**	10.49	34.08	55.42
Concreteness	11	21	0.690	0.351–1.029	<0.001***	0.394***	0.026	13.94	80.78	5.28
Bizarreness	7	11	1.078	0.742 –1.413	<0.001***	0.132*	0.000	45.62	54.38	0.00

Notes: *s* = number of studies; *k* = number of effect sizes; CI = confidence interval; Mean *g* = mean effect size (Hedge's *g*); CI = confidence interval; % Var = percentage of variance explained; σ^2 level2 = variance between effect sizes within the same study; σ^2 level3 = variance between studies.

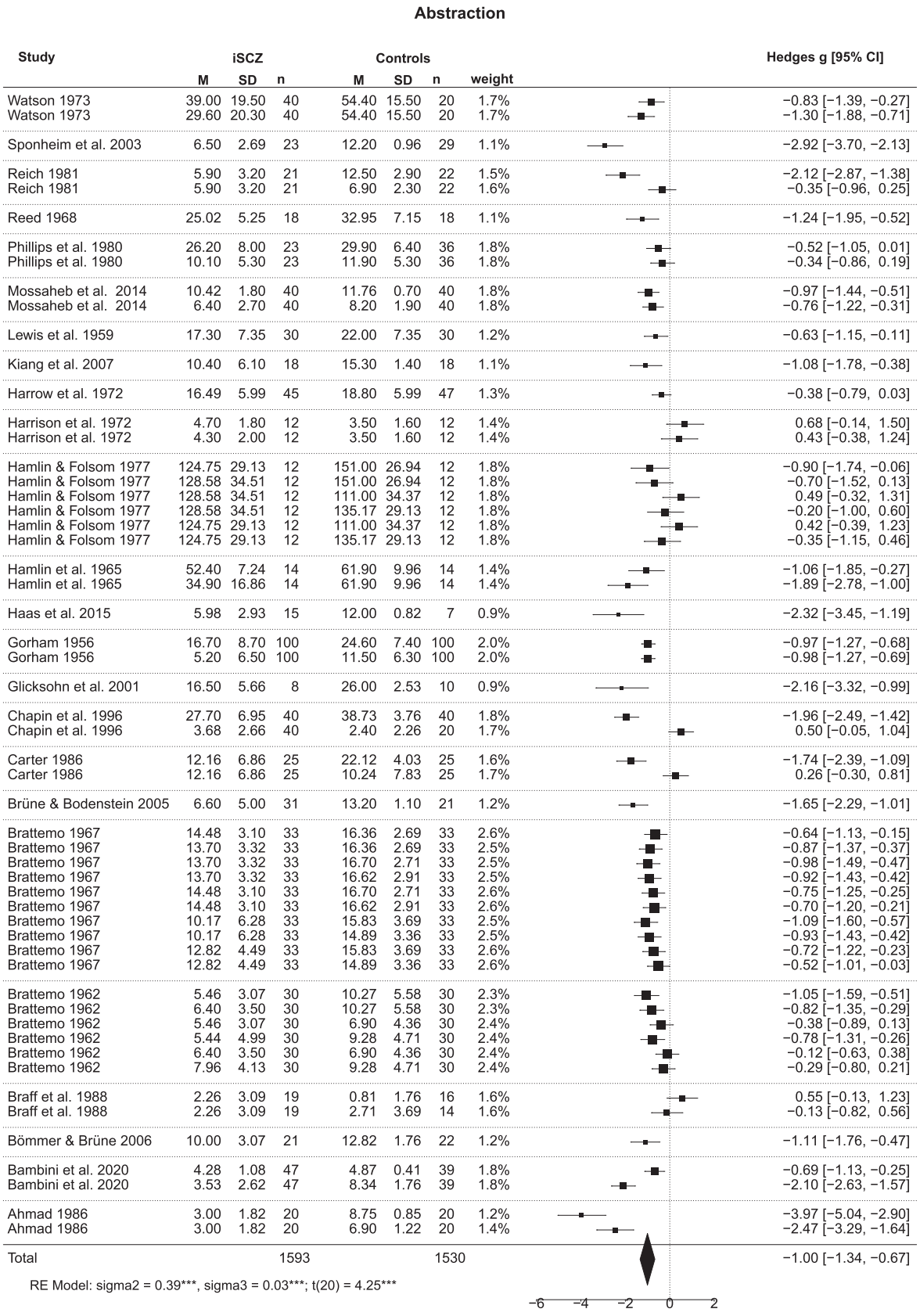


Fig. 2. Forest plot of effect sizes comparing individuals with schizophrenia (iSCZ) and controls on abstraction in their interpretation of proverbs. Dotted lines are used to delineate clusters of effect sizes within each study. Negative effect sizes indicate less abstraction in iSCZ compared to controls.
Note: M = mean, SD = standard deviation, CI = confidence interval, iSCZ = individuals with schizophrenia, sigma2 = within-study heterogeneity (level 2), sigma3 = between-study heterogeneity (level 3). * $p < .05$, *** $p < .001$.

Table 3
Moderator analyses results for abstraction in the interpretation of proverbs.

Moderators	Abstraction				
	β_0 (mean *g*)	t_0	β_1	t_1	$F(df_1, df_2)$
Type of controls					F (1, 53) = 19.392***
CC (RC)	−0.535	−2.929**			
HC	−1.312	−7.938***	−0.777	−4.407***	
Response format					F (1, 53) = 0.000
Verbal response (RC)	−1.005	−5.463***			
Multiple-choice	−1.002	−4.708***	0.004	0.018	
Proverbs from					F (5, 49) = 0.442
GPT (RC)	−1.012	−3.901***			
Blaufarb (1962)	−0.725	−1.369	0.253	0.398	
Barth & Küfferle (2001)	−1.483	−2.800**	−0.639	−1.023	
Benjamin (1944)	−0.436	−0.579	0.630	0.660	
Self-selected	−0.838	−2.257*	0.139	0.330	
Subsection of a test battery	−1.083	−1.118	−0.314	−0.465	
Scoring system					F (6, 48) = 1.027
Gorham (RC)	−1.432	−4.366***			
Barth & Küfferle (2001)	−1.647	−3.056**	−0.215	−0.341	
Blaufarb (1962)	−0.747	−1.395	0.685	1.090	
Benjamin (1944)	−1.180	−1.490	0.252	0.293	
Meadow et al. (1953)	−0.504	−0.819	0.928	1.330	
Self-developed	−1.250	−2.120*	0.817	1.893	
Language					F (5, 49) = 2.733*
English (RC)	−0.794	−4.674***			
German	−1.157	−2.798**	−0.362	−0.810	
Italian	−1.372	−2.190*	−0.577	−0.889	
Swedish	−0.697	−1.905†	0.098	0.242	
Hindi	−3.130	−4.534***	−2.335	−3.284**	
Hebrew	−2.156	−2.330*	−1.362	−1.447	
Stadium of iSCZ					F (2, 41) = 1.673
Acute (RC)	−0.616	−2.903**			
Chronic	−0.706	−3.335**	−0.090	−0.397	
Mixed	−1.163	−5.369	−0.547	−1.804†	
DSM III					F (1, 53) = 1.232
Before (RC)	−0.821	−3.655***			
After	−1.159	−5.338***	−0.338	−1.108	
Continuous Moderators[1]					
Age	0.351	3.529**			F (1, 43) = 12.454***
IQ/ Education	−0.395	−2.934**			F (1,33) = 8.609**
% Females	−0.262	−1.187			F (1,20) = 1.409

Note: β_0 = intercept/mean effect size (*g*); t_0 = difference in mean *g* with zero; β_1 = estimated regression coefficient; t_1 = difference in mean *g* with reference category; $F(df_1, df_2)$ = omnibus test, one-sided; (RC) = reference category; CC = clinical control group; HC = healthy control group.

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

[1] Difference between iSCZ and control group (iSCZ minus controls).

control group consisted of clinical individuals.

3.4. Publication bias

Inspection of the funnel plot and Eggers' test indicated a symmetric distribution of effect sizes for abstraction (bias = −0.503, [*CI*: −0.957, −0.049]; $t = -1.607$, $df = 22$, $p = .122$), concreteness (bias = 0.785, [*CI*: −0.012, 1.574], $t = -0.267$, $df = 9$, $p = .795$) and bizarreness (bias = 0.574 [CI: −0.451, 1.599], $t = 1.216$, $df = 5$, $p = .278$). However, the trim and fill method identified one missing study for bizarreness located on the left side of the plot, indicating that small and insignificant effect sizes seemed underrepresented (see Figs. 1–3 in the Supplementary Material).

4. Discussion

In the current meta-analysis, we investigated proverb comprehension in individuals with SCZ compared to clinical and healthy control groups. Our study focused on three distinct answering patterns that have been previously proposed to occur in individuals with SCZ: absence of abstraction, concrete responses, and bizarre responses. The results of our analysis revealed that, on average, individuals with SCZ exhibited lower levels of abstract interpretations, higher levels of concrete interpretations, and an increased occurrence of bizarre responses compared to HC. When comparing the effects to clinical control groups, the differences in bizarre and abstract responses were relatively smaller. Surprisingly, no significant differences were observed between individuals with SCZ and CC for concrete responses. This finding contradicts the traditional assumption that concreteness, defined as a literal response to nonliteral language [9], is specific to schizophrenia [6,8]. This also corroborates early doubts on the reliability and validity of proverbs tasks [1,19]. Especially, in cases in which concretism is defined as a continuum with concreteness and abstraction as the end-points, as is was suggested early on [14] and implicitly operationalized in studies on concretism [65,66,79]. Instead, concreteness seems transdiagnostic.

It seems that the comprehension of proverbs may rather be characterized as a continuum of abstraction and bizarre or idiosyncratic thinking. This is supported by longitudinal work, showing that the underlying structure of semantic abstraction comprised increased conceptual overinclusion and idiosyncratic associations [39]. It should be noted, however, that when assessing at the study level, our analyses showed evidence of publication bias in bizarreness, suggesting that the higher bizarreness in individuals with SCZ compared to controls may be inflated by the non-publication of studies with null results. As the number of studies assessing bizarre responses was generally low, this suggests a need for further research.

We also examined other potential moderating factors. People with SCZ showed comparable effects on proverb comprehension across different response formats (verbal and multiple-choice) and are not significantly influenced by factors such as scoring system, proverbs asked, gender, or acute or chronic stage of the disorder. For abstract responses, we found that both age and IQ/education were significant moderators, as well as a tendency for bizarreness.

This has several implications. First, it highlights that abstraction (and bizarreness) may be more biased by factors not inherently related to SCZ psychopathology (age and IQ) than concrete or bizarre responses. Cultural knowledge plays a general role in the comprehension of proverbs [80,81]; however, the results strengthen the assumption that there are other mechanisms accounting for concrete in individuals with SCZ.

Second, the moderating effect of age implied that when individuals with SCZ were older than controls, their deficit in abstract responses appeared to be greater. In general, longitudinal data has shown that abstraction capacity rather improves than decreases with age in individuals with SCZ [39]. Assuming that the discrepancy between our findings and the longitudinal data is not based on training effects, we propose that several factors related to the composition of our meta-

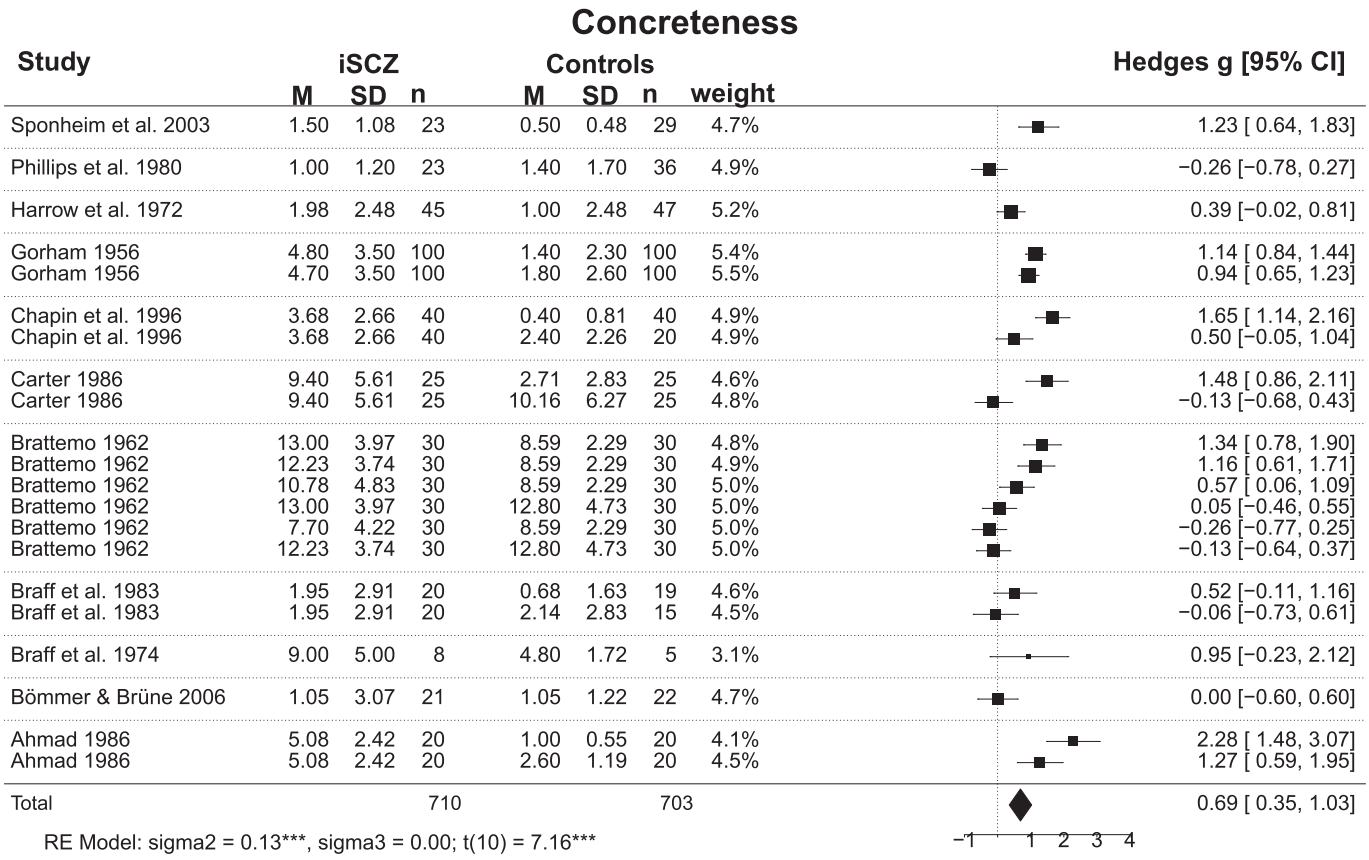


Fig. 3. Forest plot of effect sizes comparing individuals with schizophrenia (iSCZ) and controls on concreteness in their interpretation of proverbs. Dotted lines are used to delineate clusters of effect sizes within each study. Positive effect sizes indicate more concreteness in iSCZ compared to controls.
Note: M = mean, SD = standard deviation, CI = confidence interval, iSCZ = individuals with schizophrenia, sigma2 = within-study heterogeneity (level 2), sigma3 = between-study heterogeneity (level 3). *** $p < .001$.

analysis might have contributed to this contrasting effect. The majority of studies included in our meta-analysis assessed hospitalized individuals, who likely had longer and more severe disorder histories, particularly as they grew older. This prolonged illness history could have affected their cognitive functioning, leading to greater deficits in abstract responses in older individuals with SCZ compared to controls. Furthermore, it is important to consider the historical context of the studies included in our meta-analysis. The studies were predominantly conducted in the 1960s, and participants with SCZ in those studies have experienced longer durations without antipsychotic treatment.

Third, responses were also moderated by education and intelligence, such that lower IQ/education in individuals with SCZ increased group differences in abstraction and tended to increase them in bizarreness. As proverb comprehension has been shown to be related to intelligence and education [2,19,59], this may have artificially inflated group differences in some studies. It is noteworthy, however, that even in groups matched for IQ, differences in abstraction remained significant. This suggests that the deficit in abstraction observed in individuals with SCZ cannot be attributed solely to differences in intelligence between the groups.

Our results therefore do not encourage the use of concrete responses to proverbs as a good discriminant marker between schizophrenia and other diagnoses, as suggested in clinical practice. Our outcome is supported by longitudinal data following proverb interpretations in individuals with SCZ, affective psychosis, and unipolar depression over 20 years [39]. Consistent with these findings, our results imply the requirement for a more differentiated investigation of concretism, that takes not only the psychopathology of SCZ, but also the linguistic characteristics of the figure of speech into account.

4.1. *Implications for future research*

Our findings suggest several areas for future research. First, they underscore the need to implement more differentiated response classifications. Despite initial approaches [18,19,22,23], the recent research has only focused on the abstraction and concreteness [42,43]). Notably, abstraction has often been conflated with the correct interpretation of proverbs [18,42,79]. While these studies discuss their findings as a measure of concretism, lower test scores in these studies do not necessarily indicate concrete responses or a lack of abstraction. Other incorrect explanations (e.g., tangential, illogical or bizarre responses) would also score lower. To overcome this, it requires research to shift from multiple-choice to verbal response formats. Multiple-choice formats may not fully capture the reasons behind an individual's choice of an alternative answer, potentially leading to misunderstandings or overlooking other factors influencing their responses apart from lack of abstraction. Furthermore, multiple-choice is connected to executive functions. In patients with schizophrenia, executive function appears to be associated with proverb comprehension performance [17,31,59,82]. Verbal response formats, on the other hand, allow participants to provide detailed explanations for their answers, enabling researchers to gain deeper insights into the cognitive processes involved in proverb comprehension. Furthermore, it is important to recognize that the relationship between proverb comprehension and abstraction is subject to linguistic debate. Studies have shown that proverb comprehension may not solely reflect a direct expression of abstraction ability [83–85]. Hence, incorporating new developments in cognitive linguistics, such as the embodiment perspective, could offer valuable insights into the underlying mechanisms of proverb comprehension deficits in individuals

Table 4
Moderator analyses results for concreteness in the interpretation of proverbs.

Moderators	Concreteness				
	β_0 (mean *g*)	t_0	β_1	t_1	$F(df_1, df_2)$
Controls					F (1, 19) = 46.110***
CC (RC)	0.133	0.695			
HC	1.216	6.299***	1.082	6.790***	
Response format					F (1, 19) = 0.129
Verbal response (RC)	0.738	3.428**			
Multiple-choice	0.622	2.556*	−0.116	−0.360	
Proverbs from					F (2, 18) = 0.596
GPT (RC)	0.759	4.141***			
Barth & Küfferle (2001)	0.000	0.000	−0.759	−0.998	
Benjamin (1944)	0.393	0.558	−0.366	−0.503	
Scoring system					F (3, 17) = 1.609
Gorham (RC)	1.043	4.120***			
Barth & Küfferle (2001)	0.000	0.000	−1.043	−1.430	
Hertler et al. (1978)	1.232	1.801†	0.189	0.259	
Self-developed	0.470	2.376*	−0.573	0.092†	
Language					F (3, 17) = 2.176†
English (RC)	0.697	3.661**			
German	0.000	0.000	−0.697	−1.007	
Swedish	0.446	1.685	−0.251	−0.769	
Hindi	1.754	3.539**	1.057	1.991†	
Stadium of iSCZ					F (2,11) = 0.585
Acute (RC)	0.249	0.627			
Chronic	0.652	2.097	0.403	0.798	
Mixed	0.777	2.666*	0.528	1.071	
DSM III					F (1, 19) = 0.097
Before (RC)	0.635	2.438*			
After	0.743	3.221**	0.108	0.312	
Continuous Moderators[1]					
Age	−0.153	−0.764			F (1,13) = 0.584
IQ/ Education	0.266	1.440			F (1,12) = 2.075
% Females	0.451	1.548			F (1, 6) = 2.397

Note: β_0 = intercept/mean effect size (*g*); t_0 = difference in mean *g* with zero; β_1 = estimated regression coefficient; t_1 = difference in mean *g* with reference category; $F(df_1, df_2)$ = omnibus test, one-sided; (RC) = reference category; CC = clinical control group; HC = healthy control group.

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

[1] Difference between iSCZ and control group (iSCZ minus controls).

with SCZ [86,87].

The fact that over 90% of the clinical controls in the current meta-analysis included individuals with depressive disorders further strengthens the need for future studies to examine whether that concrete responses may be specifically related to aspects incorporated in negative symptoms, that may overlap between diagnoses. In addition to better assessments, future work is needed to understand the role of positive and negative symptoms in relation to *type* of response. For example, negative and positive symptoms may lead to a miscomprehension of proverbs in different manners. Individuals with negative symptoms suffer from a narrowed range of experience (i.e., narrowed association; [88]), which may promote them to choose only semantically close meanings when looking for an interpretation. Indeed, incorrect interpretations of metaphors been associated with negative symptoms [42,89] but not with thought disorder [89]. This is further supported by research on individuals with depression, who tend to give concrete answers [29,33]. In the current analysis, insufficient data were available to conduct subgroup analyses on positive and negative symptoms. This scarcity of data is, in part, because a significant proportion of studies included in the meta-analysis were conducted before the implementation of the positive and negative symptoms dichotomy [90]. From the available studies that did assess these symptoms, the prevailing evidence indicates no correlation between positive symptoms and abstraction scores ([31,43,59,82] but see [91]). However, tasks requiring active verbalization or paraphrasing appear to be specifically impacted by negative symptoms [42]. Notably, only one study has probed the relationship between positive and negative symptoms in connection with bizarre and concrete responses [31]. Their findings show a lack of association with concrete responses but a significant correlation with bizarre responses. This underscores the pressing need for more extensive research to elucidate the intricate relationships between these symptoms and response types.

In contrast to the prevailing view that loosened associations in individuals with SCZ primarily lead to concrete interpretations, our proposal posits that these associations may be more related to bizarre and idiosyncratic responses. Bleuler's seminal work on schizophrenia established that individuals with this disorder experience disruptions in the connections between mental associations, resulting in unusual and logically false interpretations, often associated with positive symptoms such as thought disorder. We argue that when differentiating between type of proverb interpretations, a loosening of associations or positive symptoms observed in individuals with SCZ may contribute more prominently to bizarre than literal-concrete responses. Bizarre and idiosyncratic interpretations occur when the associated meanings are disconnected from the literal meaning, leading to illogical and peculiar responses. Importantly, our findings indicate that while concrete responses did not differentiate between the predominantly depressive clinical controls and individuals with schizophrenia, bizarre-idiosyncratic responses did. Although caution must be exercised in interpreting these results due to the small number of studies, potential author biases (e.g. the seminal work by Braff [50,51,62] and Brattemo [57,30]) and lack of comprehensive matching in some cases, they do allow for a possible future direction of research. In support of our proposal and the need for verbal accounts, a recent study by Srivastava, Bilgrami [92] utilized natural language processing to analyze verbal accounts of individuals with schizophrenia. They demonstrated that individuals at high risk for psychosis used more metaphorical language than healthy controls, but only the amount of bizarre language was indicative of a clinical diagnosis of SCZ. Therefore, focusing on bizarre-idiosyncratic responses in individuals with SCZ may aid in establishing proverb interpretation as a more indicative measure for individuals with or at risk for SCZ [23–25,88].

Lastly, none of the studies controlled for the most intuitive confounding variable: individual familiarity with the stimulus. Proverbs are highly conventionalized and convey conserved social and cultural knowledge [81,93]. Given that unknown proverbs are difficult for healthy individuals as well [94–96], it remains unclear whether studies simply assessed a lack of semantic knowledge in individuals with SCZ. Of the included studies, only Mossaheb, Aschauer [42] compared novel and conventional expressions. They found that patients scored lower on both conventional and novel metaphoric proverbs. However, in their study, and many other experimental designs, the classification of proverbs as familiar or unfamiliar is typically determined by researchers, independent raters, or dictionaries. Yet, participants' actual familiarity with proverbs could differ from researchers' or raters' judgments. This introduces a confounding variable that is not necessarily captured by IQ and education and cannot be eliminated by matching. To date, two studies have demonstrated that individuals with SCZ are subjectively less familiar with figurative expressions than controls [17,97]. Rapp, Felsenheimer [97] found that the alleged deficit in the comprehension of conventional metaphors disappeared when only familiar metaphors

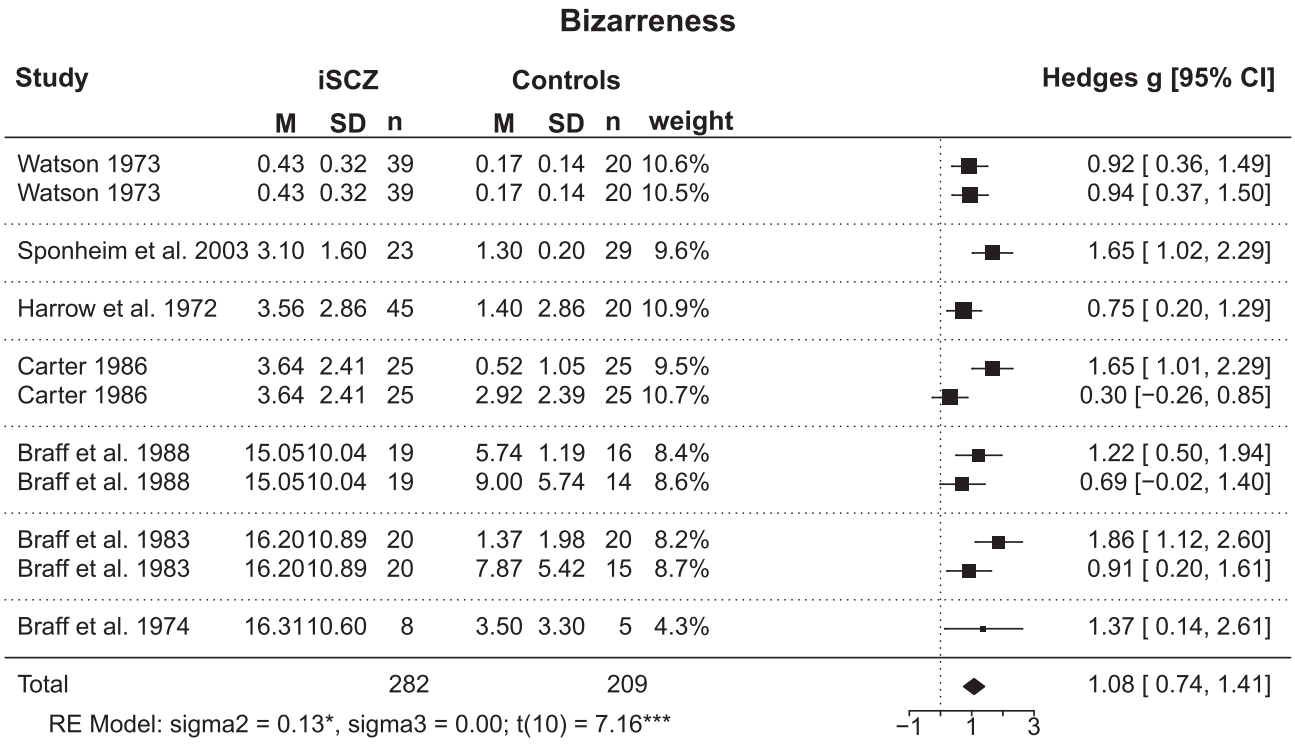


Fig. 4. Forest plot of effect sizes comparing individuals with schizophrenia (iSCZ) and controls on bizarreness in their interpretation of proverbs. Dotted lines are used to delineate clusters of effect sizes within each study. Positive effect sizes indicate more bizarreness in iSCZ compared to controls.
Note: M = mean, SD = standard deviation, CI = confidence interval, iSCZ = individuals with schizophrenia, sigma2 = within-study heterogeneity (level 2), sigma3 = between-study heterogeneity (level 3). *$p < .05$, *** $p < .001$.

were considered, whereas Thoma, Hennecke [17] found that differences in understanding persisted. Although familiarity does not seem to explain the whole deficits, these results indicate that familiarity clearly affects alleged group differences.

From a broader perspective, the cultural knowledge required for the interpretation of proverbs may contribute to misdiagnosis in individuals with different ethnocultural backgrounds [98]. For example, research has shown that individuals from African American backgrounds demonstrate greater recognition of proverbs drawn from African American culture compared to individuals from Caucasian backgrounds [99]. In addition, the GPT is related to socioeconomic background [95]. In line with this, participants with and without a risk for psychosis become more literal in their interpretations when confronted with unfamiliar proverbs [88].

In this sense, language emerged in our meta-analysis as a significant moderator for abstraction and tendency for concreteness. The majority of studies were conducted in English, limiting the generalizability of the results across different languages and cultures. Of the included studies, only two analyzed non-Germanic or non-Romance languages [61,79]. The cultural habits of figurative language may differ between different language communities [100], but the current scientific debate on concretism lacks interlinguistic comparisons [98]. Proverbs cannot be translated one-to-one [96,101]; therefore, understanding the extent to which linguistic/cultural influences shape alleged pragmatic language impairments in SCZ requires extending research on figurative language to cross-linguistic comparisons.

4.2. Limitations

Our study acknowledges several inherent limitations that must be considered when interpreting the findings. First, diagnostic criteria for SCZ have evolved over the decades. This evolution has led to variations in the definition and categorization, introducing potential confounding factors. For example, the lack of moderation by acute or chronic stage indicated that proverb comprehension deficits in individuals with SCZ were consistent across different stages of the illness. However, earlier studies often included individuals with diverse disease stages and characteristics, with some studies even encompassing patients who had undergone lobotomy procedures [30]. Another complexity arises from the historical classification of symptom severity, which has used different frameworks, some of which are no longer used in contemporary research. For instance, the works of Kantor, Wallner [60] introduced distinctions between process and reactive symptomatology, a categorization that is now unfamiliar. This heterogeneity in disorder classification and symptom severity assessment warrants caution when interpreting our results and may limit our ability to establish a clear moderating effect based on the acute or chronic stage of individuals with SCZ.

The general limitations of meta-analyses also apply to our study. Some studies have small numbers of cases. Important confounders may have been underestimated because they were not consistently reported (see Table 1). Heterogeneity was very high in some of the analyses, although our procedure allowed us to distinguish that heterogeneity for concreteness and bizarreness was mostly due to within-study factors, whereas for abstraction both within-study and between-study factors contributed. Linguistic factors also limit the validity of our meta-analysis. Proverbs are not consistently defined in scientific studies [102,103]. Proverbs are not equally intelligible, and comprehensibility varies with: literality, the extent to which the idiom is literally plausible; compositionality, the extent to which the meanings of individual words in the proverbs contribute to its figurative meaning; and the presence of contextual bias [104,105]. We did not include other types of fixed verbal expressions and figurative language (e.g., metaphors or irony and sarcasm), which are related to specific neural correlates in comprehension (see [45,106]) and are possibly associated with other psychopathological dimensions like social cognition [107,108].

Moreover, we have severely limited our conclusions by not including the studies using the PANSS [109] in this paper. The PANSS includes an

Table 5
Moderator analyses results for bizarreness in the interpretation of proverbs.

Moderators	Bizarreness				
	β_0 (mean *g*)	t_0	β_1	t_1	$F(df_1, df_2)$
Controls					F (1, 9) =
CC (RC)	0.651	3.561**			11.259**
HC	1.409	8.335***	−0.758	−3.356**	
Proverbs from					F (1, 9) =
GPT (RC)	1.120	6.822***			0.558
Benjamin (1944)	0.746	1.581	−0.373	−0.747	
Scoring system					F (2, 8) =
Shimkunas (1967) (RC)	1.098	5.177***			0.269
Benjamin (1944)	0.746	1.480	−0.352	−0.643	
Marengo et al. (1986)	1.172	3.875**	0.074	0.201	
Language					F (1, 9) =
English (RC)	1.120	6.822***			0.558
Swedish	0.746	1.581	−0.373	−0.747	
Stadium of iSCZ					F (2, 4) =
Acute (RC)	0.936	1.567			0.172
Chronic	0.941	2.701†	0.005	0.007	
Mixed	1.223	3.234*	0.287	0.406	
DSM III					F (1, 9) =
Before (RC)	0.928	3.639**			0.546
After	1.164	6.059***	0.236	0.739	
Continuous Moderators[1]					
Age	−0.310	−1.883†			F (1, 8) = 3.547*
IQ/Education	0.300	1.910†			F (1, 7) = 3.649*
% Females	0.074	0.283			F (1, 4) = 0.080

Note: β_0 = intercept/mean effect size (*g*); t_0 = difference in mean *g* with zero; β_1 = estimated regression coefficient; t_1 = difference in mean *g* with reference category; $F(df_1, df_2)$ = omnibus test, one-sided; (RC) = reference category; CC = clinical control group; HC = healthy control group.

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

[1] Difference between iSCZ and control group (iSCZ minus controls).

item "Difficulty with Abstract Thinking" (item N5), which comprises (but is not limited to) reasoning about the meaning of proverbs. The rating instructions for this item show very high parallels with the free speech responses in the proverb studies reported here (see appendix in [110]). However, the 7-point Likert scale of the rating instructions does not rigorously differentiate between abstract, concrete, or bizarre-idiosyncratic responses, merging them all into one scale [109]. Because of the extreme prevalence of the PANSS and its well-studied relationship to, e.g., the course and therapeutic responsiveness of psychosis, the analysis of this item allows for an analysis of proverb explanation on a much broader basis than the experimental studies included here. A meta-analysis specifically on PANSS N5 and proverbs would be useful. Nevertheless, the main conclusions here apply in a similar way to the analysis of the PANSS abstract thinking item. For example, the PANSS also scores bizarre-idiosyncratic responses higher than single concrete responses. Furthermore, the risk of overestimation due to socioeconomic background is similar [98].

Although not the focus of this paper, imaging studies of proverbs in schizophrenia would also be worth considering. How proverb comprehension relates to executive functioning would be interesting to explore further [17], both clinically and neurobiologically. There is a key role of the left inferior frontal gyrus for proverb comprehension, therefore hypofrontality in schizophrenia may play a role [111,112]. However, study results are mixed [112–115], with some studies even showing increased left prefrontal activation in schizophrenia [113].

As the original project started before the development of pre-registrations as a common practice in research this meta-analysis [116], could not be pre-registered. To address this, we have highlighted which changes in the statistical process were modified and which moderators were unplanned and therefore have to be considered exploratory. The extracted research data are available online (https://osf.io/f2qpc/).

4.3. Conclusions

The qualitative details of individuals with SCZs' thoughts and perceptions involved in proverb comprehension remains unclear after 70 years of research. Given the decade-long role of proverb interpretation in regular psychopathology assessments, we conclude that its diagnostic specificity on the basis of the existing studies is astonishingly low. Individuals with SCZ could not be differentiated from CC in terms of the most acknowledged concept of concretism, concrete or literal responses, which is often considered the opposite of abstraction.

As clinical and scientific importance of proverbs (and figurative language in general) is increasingly recognized in individuals with dementia [117] and autism [118], these results underline significant research gaps with a need for future research. Due to a lack of studies, we were not yet able to conduct sub-analyses and comparisons between these disorders and individuals with SCZ, which would be desirable. The same is true for other disorders such as bipolar disorder, delirium, borderline personality disorder and attention deficit hyperactivity disorder.

The current meta-analyses revealed a strong need for a unifying definition of what is (and what is not) understood as "concretism" to guarantee unbiased diagnostic discrimination. These results refute the notion that concretism is only a characteristic of individuals with SCZ and is fully explained by a lack of abstraction. A promising discriminatory factor may be a bizarre-idiosyncratic responses, although the sparsity of studies impedes drawing a definite conclusion at present. Whether the degree and type of symptomatology (negative vs. positive symptoms), rather than the diagnosis, may be a potential predictor of the underlying causes of proverb misinterpretation requires further investigation. Detecting bizarre-idiosyncratic responses relies on analyzing free speech, but a decline in qualitative language data has impeded this line of research in recent decades. Ultimately, advancing proverb comprehension research in schizophrenia will not only inform clinical assessments but also offer valuable insights into the cognitive complexities of this disorder. Notably, this pursuit also underscores the significance of incorporating considerations for cultural and social backgrounds, aspects that have often been overlooked in previous studies. By refining our understanding of proverb comprehension deficits and considering cultural diversity, we can enhance the diagnostic accuracy for individuals with SCZ, leading to more effective clinical evaluations and appropriate interventions tailored to their specific challenges.

CRediT authorship contribution statement

Anne Felsenheimer: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Alexander M. Rapp:** Conceptualization, Data curation, Investigation, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comppsych.2023.152444.

References

[1] Andreasen NC. Reliability and validity of proverb interpretation to assess mental status. Compr Psychiatry 1977;18:465–72.
[2] Gorham DR. Use of the proverbs test for differentiating schizophrenics from normals. J Consult Psychol 1956;20:435–40.
[3] Mitchell RL, Crow TJ. Right hemisphere language functions and schizophrenia: the forgotten hemisphere? Brain. 2005;128:963–78.
[4] Kleist K. Aphasie und Geisteskrankheit. Münch Med Wochenschr 1914;61:8–12.
[5] Bleuler E. Dementia praecox oder Gruppe der Schizophrenien. 1st ed. Leipzig [u. a.]: Deuticke; 1911.
[6] Blaufarb H. A demonstration of verbal abstracting ability in chronic schizophrenics under enriched stimulus and instructional conditions. J Consult Psychol 1962;26:471–5.
[7] Chapman LJ. Confusion of figurative and literal usages of words by schizophrenics and brain damaged patients. J Abnorm Soc Psychol 1960;60:412–6.
[8] Kasanin JS. Language and thought in schizophrenia. Berkeley: University of California Press; 1944.
[9] Goldstein K. Methodological approach to the study of schizophrenic thought disorder. In: Kasanin JS, editor. Language and thought in schizophrenia : Collected papers presented at the meeting of the American Psychiatric Association, May 12, 1939, Chicago, Illinois. New York: Norton / University of California Press; 1944.
[10] Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13:261–76.
[11] Lindenmayer JP, Harvey PD, Khan A, Kirkpatrick B. Schizophrenia: measurements of psychopathology. Psychiatr Clin North Am 2007;30:339–63.
[12] Sadock BJ, Sadock VA, Ruiz P. Comprehensive textbook of psychiatry. 50th Anniversary Edition ed. Philadelphia: lippincott Williams & Wilkins Philadelphia; 2000.
[13] Benjamin JD. A method of distinguishing and evaluating formal thinking disorder. In: Kasanin JS, editor. Language and thought in schizophrenia. Berkeley, CA, US: University of California Press; 1944. p. xiv, 133-xiv.
[14] Gorham DR. Additional norms and scoring suggestions for the Proverbs Test. Psychol Rep 1963;13:487–92.
[15] Barth A, Kufferle B. Die Entwicklung eines Sprichworttests zur Erfassung konkretistischer Denkstörungen bei schizophrenen Patienten. Nervenarzt. 2001;72:853–8.
[16] Delis DC, Kaplan E, Kramer JH. The Delis-Kaplan executive function system: Technical manual. San Antonio: The Psychological Corporation. Pearson Education Limited; 2001.
[17] Thoma P, Hennecke M, Mandok T, Wähner A, Brüne M, Juckel G, et al. Proverb comprehension impairments in schizophrenia are related to executive dysfunction. Psychiatry Res 2009;170:132–9.
[18] Meadow A, Greenblatt M, Solomon HC. "Looseness of association" and impairment in abstraction in schizophrenia. J Nerv Ment Dis 1953;118:27–35.
[19] Shimkunas AM, Gynther MD, Smith K. Abstracting ability of schizophrenics before and during phenothiazine therapy. Arch Gen Psychiatry 1966;14:79–83.
[20] Hertler CA, Chapman LJ, Chapman JP. A scoring manual for literalness in proverb interpretation. J Consult Clin Psychol 1978;46:551–5.
[21] Watson CG, Plemel D, Burke M. Proverb test deficit in schizophrenia and brain-damaged patients. J Nerv Ment Dis 1979;167:561–5.
[22] Marengo JT, Harrow M, Lanin-Kettering I, Wilson A. Evaluating bizarre-idiosyncratic thinking: a comprehensive index of positive thought disorder. Schizophr Bull 1986;12:497–511.
[23] Harrow M, Tucker GJ, Adler D. Concrete and idiosyncratic thinking in acute schizophrenic patients. Arch Gen Psychiatry 1972;26:433–9.
[24] Adler D, Harrow M. Idiosyncratic thinking and personally overinvolved thinking in schizophrenic patients during partial recovery. Compr Psychiatry 1974;15:57–67.
[25] Carpenter BN, Chapman LJ. Premorbid status in schizophrenia and abstract, literal or autistic proverb interpretation. J Abnorm Psychol 1982;91:151–6.
[26] Reich JH. Proverbs and the modern mental status exam. Compr Psychiatry 1981;22:528–31.
[27] Swinney DA, Cutler A. The access and processing of idiomatic expressions. J Verbal Learn Verbal Behav 1979;18:523–34.
[28] Colston HL, Athanasiadou A. Irony in language use and communication. Amsterdam: John Benjamins Publishing Company; 2017.
[29] Jung-Beeman M. Bilateral brain processes for comprehending natural language. Trends Cogn Sci 2005;9:512–8.
[30] Brattemo C. Interpretations of proverbs in schizophrenic patients. Furth Stud Acta Psychol 1962;20:254–63.
[31] Sponheim SR, Surerus-Johnson C, Leskela J, Dieperink ME. Proverb interpretation in schizophrenia: the significance of symptomatology and cognitive processes. Schizophr Res 2003;65:117–23.
[32] Bell NL, Rucker M, Finch AJ, Alexander J. Concurrent validity of the Slosson full-range intelligence test: comparison with the Wechsler intelligence scale for children-third edition and the woodcock Johnson tests of achievement-revised. Psychol Sch 2002;39:31–8.
[33] Uekermann J, Thoma P, Daum I. Proverb interpretation changes in aging. Brain Cogn 2008;67:51–7.
[34] Szepietowska EM, Filipiak S. Interpretation of familiar metaphors and proverbs by polish people in middle and late adulthood. Int J Lang Commun Disord 2021;56:841–57.
[35] Harrison AW, Spellman MS, Mellsop GW. The proverbs test for disorder of thinking in schizophrenia and mania. Aust N Z J Psychiatry 1972;6:52–6.
[36] Hamlin RM, Haywood HC, Folsom AT. Effect of enriched input on schizophrenic abstraction. J Abnorm Psychol 1965;70:390–4.
[37] Watson CG. Abstract thinking deficit and autism in process and reactive schizophrenics. J Abnorm Psychol 1973;82:399–403.
[38] Hamlin RM, Folsom AT. Impairment in abstract responses of schizophrenics, neurotics, and brain-damaged patients. J Abnorm Psychol 1977;86:483–91.
[39] Rosen C, Harrow M, Tong L, Jobe TH, Harrow H. A word is worth a thousand pictures: a 20-year comparative analysis of aberrant abstraction in schizophrenia, affective psychosis, and non-psychotic depression. Schizophr Res 2021;238:1–9.
[40] Parola A, Brasso C, Morese R, Rocca P, Bosco FM. Understanding communicative intentions in schizophrenia using an error analysis approach. NPJ Schizophr 2021:7.
[41] Haas MH, Chance SA, Cram DF, Crow TJ, Luc A, Hage S. Evidence of pragmatic impairments in speech and proverb interpretation in schizophrenia. J Psycholinguist Res 2015;44:469–83.
[42] Mossaheb N, Aschauer HN, Stoettner S, Schmoeger M, Pils N, Raab M, et al. Comprehension of metaphors in patients with schizophrenia-spectrum disorders. Compr Psychiatry 2014;55:928–37.
[43] Bambini V, Arcara G, Bosinelli F, Buonocore M, Bechi M, Cavallaro R, et al. A leopard cannot change its spots: a novel pragmatic account of concretism in schizophrenia. Neuropsychologia. 2020;139:107332.
[44] Rapp A. The role of the right hemisphere for language in schizophrenia. In: Sommer IK, Kahn RS, editors. Language lateralization and psychosis. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, Dehli: Cambridge University Press; 2009. p. 147–56.
[45] Rapp AM, Mutschler DE, Erb M. Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. Neuroimage. 2012;63:600–10.
[46] Rapp AM. Comprehension of metaphors and idioms: an updated meta-analysis of functional magnetic resonance imaging studies. In: de Zubicaray GI, Schiller NO, editors. The Oxford handbook of neurolinguistics. Oxford: Oxford University Press; 2019. p. 710–35.
[47] Winner E, Gardner H. The comprehension of metaphor in brain-damaged patients. Brain. 1977;100:717–29.
[48] Reed JL. The proverbs test in schizophrenia. Br J Psychiatry 1968;114:317–21.
[49] Rapp AM, Wild B. Nonliteral language in Alzheimer dementia: a review. J Int Neuropsychol Soc 2011;17:207–18.
[50] Braff DL, Glick ID, Griffin P. Thought disorder and depression in psychiatric patients. Compr Psychiatry 1983;24:57–64.
[51] Braff DL, Glick ID, Johnson MH, Zisook S. The clinical significance of thought disorder across time in psychiatric patients. J Nerv Ment Dis 1988;176:213–20.
[52] Thoma P, Daum I. Neurocognitive mechanisms of figurative language processing—evidence from clinical dysfunctions. Neurosci Biobehav Rev 2006;30:1182–205.
[53] Rapp A, Schmierer P. Proverbs and nonliteral language in schizophrenia: a systematic methodological review of all studies published 1931-2010. Schizophr Res 2010;117:422.
[54] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.
[55] Gorham DR. A proverbs test for clinical and experimental use. Psychol Rep 1956;2:1–12.
[56] Elmore CM, Gorham DR. Measuring the impairment of the abstracting function with the proverbs test. J Clin Psychol 1957;13:263–6.
[57] Brattemo C. Interpretations of proverbs in schizophrenic and depressive patients. Acta Psychol (Amst) 1961;18:342–50.
[58] Carter ML. The assessment of thought deficit in psychotic unipolar depression and chronic paranoid schizophrenia. J Nerv Ment Dis 1986;174:336–41.
[59] Brüne M, Bodenstein L. Proverb comprehension reconsidered – 'theory of mind' and the pragmatic use of language in schizophrenia. Schizophr Res 2005;75:233–9.
[60] Kantor RE, Wallner JM, Winder CL. Process and reactive schizophrenia. J Consult Psychol 1953;17:157–62.
[61] Ahmad SI. Thinking disorder among depressive and schizophrenic patients. J Psychol Res 1986;30:8–11.
[62] Braff DL, Beck AT. Thinking disorder in depression. Arch Gen Psychiatry 1974;31:456–9.
[63] Chapin KJ, Rosenbaum G, Fields RB, Wightman LH. Multiple deficit theory of schizophrenia: incidence of markers vs symptoms. J Clin Psychol 1996;52:109–23.
[64] Phillips WM, Phillips AM, Shearn CR. Objective assessment of schizophrenic thinking. J Clin Psychol 1980;36:79–89.
[65] Reed JL. Proverbs test in schizophrenia. Br J Psychiatry 1968;114:317.
[66] Arcara G, Bambini V. A test for the Assessment of Pragmatic Abilities and Cognitive Substrates (APACS): normative data and psychometric properties. Front Psychol 2016:7.
[67] Gorham DR. Clinical manual for the proverbs test. Missoula, Montana: Psychological Test Specialists; 1956.
[68] Lewis JM, Griffith EC, Riedel AF, Simmons BA. Studies in abstraction: schizophrenia and orality: preliminary results. J Nerv Ment Dis 1959;129:564–7.
[69] Brattemo C-E. Interpretations of proverbs in schizophrenic patients. Further studies. Acta Psychol (Amst) 1962;20:254–63.
[70] Shimkunas AM, Gynther MD, Smith K. Schizophrenic responses to proverbs test – abstract concrete or autistic. J Abnorm Psychol 1967;72:128.

[71] Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;36:1–48.
[72] Assink M, Wibbelink CJM. Fitting three-level meta-analytic models in R: a step-by-step tutorial. TQMP. 2016;12:154–74.
[73] Hedges LV. Estimation of effect size from a series of independent experiments. Psychol Bull 1982;92:490–9.
[74] Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med 2003;22:2693–710.
[75] Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. Br J Psychiatry 1974;125:341–7.
[76] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Bmj. 1997;315:629–34.
[77] Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56:455–63.
[78] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
[79] Glicksohn J, Alon A, Perlmutter A, Purisman R. Symbolic and syncretic cognition among schizophrenics and visual artists. Creat Res J 2001;13:133–43.
[80] Colston HL, Katz AN. Figurative language comprehension: Social and cultural influences. Routledge; 2004.
[81] Gibbs RW, Beitel D. What proverb understanding reveals about how people think. Psychol Bull 1995;118:133–54.
[82] Kiang M, Light GA, Prugh J, Coulson S, Braff DL, Kutas M. Cognitive, neurophysiological, and functional correlates of proverb interpretation abnormalities in schizophrenia. J Int Neuropsychol Soc 2007;13:653–63.
[83] Katz AN, Cacciari C, Gibbs Jr RW, Turner M. Figurative language and thought. Oxford, New York: Oxford University Press; 1998.
[84] Kessler R, Weber A, Friedrich CK. Activation of literal word meanings in idioms: evidence from eye-tracking and ERP experiments. Lang Speech 2021;64:594–624.
[85] Conklin K, Schmitt N. Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? Appl Linguis 2008;29:72–89.
[86] Park S, Baxter T. Schizophrenia in the flesh: revisiting schizophrenia as a disorder of the bodily self. Schizophr Res 2022;242:113–7.
[87] Boulenger V, Hauk O, Pulvermueller F. Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. Cereb Cortex 2009;19:1905–14.
[88] Allen J, Schuldberg D. Positive thought disorder in a hypothetically psychosis-prone population. J Abnorm Psychol 1989;98:491–4.
[89] Langdon R, Davies M, Coltheart M. Understanding minds and understanding communicated meanings in schizophrenia. Mind Lang 2002;17:68–104.
[90] Andreasen NC, Nopoulos P, Schultz S, Miller D, Gupta S, Swayze V, et al. Positive and negative symptoms of schizophrenia: past, present, and future. Acta Psychiatr Scand 1994;90:51–9.
[91] Bömmer I, Brüne M. Social cognition in “pure” delusional disorder. Cogn Neuropsychiatry 2006;11:493–503.
[92] Srivastava A, Bilgrami Z, Garg S, Corcoran C, Cecchi G, Nelson B, et al. Increased metaphor production in open-ended speech samples of patients with prodromal and developed schizophrenia detected with NLP. Biol Psychiatry 2022;91:S50.
[93] Giora R. On our mind: Salience, context, and figurative language. New York: Oxford University Press; 2003.
[94] Cunningham DM, Ridley SE, Campbell A. Relationship between proverb familiarity and proverb interpretation: implications for clinical practice. Psychol Rep 1987;60:895–8.
[95] Penn NE, Jacob TC, Brown M. Familiarity with proverbs and performance of a black population on Gorham’s proverbs test. Percept Mot Skills 1988;66:847–54.
[96] Cieślicka AB. Bilingual figurative language processing. In: Ardila A, Cieślicka AB, Heredia RR, Rosselli M, editors. Psychology of bilingualism: The cognitive and emotional world of bilinguals. Cham: Springer International Publishing; 2017. p. 75–118.
[97] Rapp A, Felsenheimer AK, Langohr K, Klupp M. The comprehension of familiar and novel metaphoric meanings in schizophrenia: a pilot study. Front Psychol 2018:8.
[98] Berg AO, Melle I, Zuber V, Simonsen C, Nerhus M, Ueland T, et al. Modelling difficulties in abstract thinking in psychosis: the importance of socio-developmental background. Cogn Neuropsychiatry 2017;22:39–52.
[99] Brown JE, Wright-Harp WY. Cultural and generational factors influencing proverb recognition. Contemp Issues Commun Sci Disord 2011;38:26–35.
[100] Zhou P, Critchley H, Garfinkel S, Gao Y. The conceptualization of emotions across cultures: a model based on interoceptive neuroscience. Neurosci Biobehav Rev 2021;125:314–27.
[101] Cieslicka AB, Heredia RR, Garcia T. Task effects in bilingual idiom comprehension. Poznan Stud Contemp Linguis 2017;53:95–117.
[102] Hatipoglu C, Can Daskin N. A proverb in need is a proverb indeed: proverbs, textbooks and communicative language ability. South African. J Educ 2020:40.
[103] Mieder W. Proverbs: A handbook. Westport, Connecticut, London: Greenwood Press; 2004.
[104] Huber-Okrainec J, Blaser SE, Dennis M. Idiom comprehension deficits in relation to corpus callosum agenesis and hypoplasia in children with spina bifida meningomyelocele. Brain Lang 2005;93:349–68.
[105] Titone DA, Connine CM. Comprehension of idiomatic expressions - effects of PREDICTABILITY and LITERALITY. J Exp Psychol-Learn Mem Cogn 1994;20:1126–38.
[106] Hauptman M, Blank I, Fedorenko E. Non-literal language processing is jointly supported by the language and Theory of Mind networks: Evidence from a novel meta-analytic fMRI approach. Cortex 2023;162:96–114.
[107] Langdon R, Coltheart M, Ward PB, Catts SV. Disturbed communication in schizophrenia: the role of poor pragmatics and poor mind-reading. Psychol Med 2002;32:1273–84.
[108] Rocca P, Galderisi S, Rossi A, Bertolino A, Rucci P, Gibertoni D, et al. Social cognition in people with schizophrenia: a cluster-analytic approach. Psychol Med 2016;46:2717–29.
[109] Kay SR. Positive and negative syndrome scale (PANSS): Technical manual: Multi-health systems. 2006.
[110] Leontieva L, Carey K, Meszaros Z, Batki S. Development of a scoring guide for the positive and negative syndrome scale (PANSS) abstract thinking subscale. Clin Psychiatry 2019;5:61.
[111] Tran BX, Nguyen TT, Boyer L, Fond G, Auquier P, Nguyen HSA, et al. Differentiating people with schizophrenia from healthy controls in a developing country: an evaluation of portable functional near infrared spectroscopy (fNIRS) as an adjunct diagnostic tool. Frontiers. Psychiatry. 2023:14.
[112] Kircher TT, Leube DT, Erb M, Grodd W, Rapp AM. Neural correlates of metaphor processing in schizophrenia. Neuroimage. 2007;34:281–9.
[113] Mashal N, Vishne T, Laor N, Titone D. Enhanced left frontal involvement during novel metaphor comprehension in schizophrenia: evidence from functional neuroimaging. Brain Lang 2013;124:66–74.
[114] Mashal N, Vishne T, Laor N. The role of the precuneus in metaphor comprehension: evidence from an fMRI study in people with schizophrenia and healthy participants. Front Hum Neurosci 2014;8.
[115] Rapp AM, Erb M, Langohr K, Markert K. Neural correlates of metonymy comprehension in schizophrenia. Schizophr Bull 2011;37:150.
[116] Simmons JP, Nelson LD, Simonsohn U. Pre-registration: why and how. J Consum Psychol 2020;31:151–62.
[117] Chakrabarty M, Klooster N, Biswas A, Chatterjee A. The scope of using pragmatic language tests for early detection of dementia: a systematic review of investigations using figurative language. Alzheimers Dement 2023:n/a.
[118] Kalandadze T, Norbury C, Narland T, Nass K-AB. Figurative language comprehension in individuals with autism spectrum disorder: a meta-analytic review. Autism 2016;22:99–117.