AN EXPLORATION OF THE RELATIONSHIP BETWEEN FIGURATIVE LANGUAGE INTERPRETATION AND COGNITIVE ABILITIES

A Dissertation by

Ryley Skinner

Master of Arts, Wichita State University, 2020

Bachelor of Arts & Bachelor of Health Sciences, University of Missouri - Columbia, 2018

Submitted to the Department of Psychology
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

July 2023

AN EXPLORATION OF THE RELATIONSHIP BETWEEN FIGURATIVE LANGUAGE INTERPRETATION AND COGNITIVE ABILITIES

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Psychology.

_____
C. Brendan Clark, Committee Chair

_____
Rhonda Lewis, Committee Member

_____
Samantha Slade, Committee Member

_____
Carryl Baldwin, Committee Member

_____
Michael Birzer, Committee Member

Accepted for the College of Liberal Arts and Sciences

_____
Andrew Hippisley, Dean

Accepted for the Graduate School

_____
Coleen R. Pugh, Dean

ACKNOWLEDGEMENTS

This dissertation is dedicated to all the people (and animals) that believed in me and came along this graduate school journey with me, and especially for my family, Thomas and Blaise. I would not be where I am today if it weren't for you all and your continuous support.

To Dr. Brendan Clark, thank you for welcoming me into your lab and giving me the opportunity to be where I am now. Your support, guidance, and encouragement throughout graduate school shaped me into the clinician and researcher that I am today, and ultimately helped me find and pursue my passion in psychology—neuropsychology. I would also like thank each of my committee members for their time, knowledge, and patience: Samantha Slade, Rhonda Lewis, Carryl Baldwin, and Michael Birzer.

To all of my mentors and supervisors throughout my clinical work, Drs. Christina Marvin, Rachel Petts, Ryan Schroeder, Phillip Martin, Selena Jackson, and Marci Young, along with many other senior graduate students and postdocs. It was through your guidance that I was able to find what I truly loved in psychology and curate my own clinical style. Thank you for the many years of training and supervision that has propelled me to my future.

To my research assistants, Hannah Piros and Taylor Keeling, thank you for all of your time and dedication to this project. You volunteered numerous hours across two semesters to learn about this project and collect, score, and input data. I could not have completed this project this year without your dedication to this project.

To my family, Thomas, Blaise, and our many animals—thank you. You all bore the brunt of the many evening and weekend sacrifices that were required to finish this project. Your love, support, flexibility, and willingness to go along this journey with me are what made any of this possible. I am so grateful that you were alongside me in this chapter of life, and I am excited to

finally be closing this chapter together. I can't wait to see what the next one holds! You all have made the long hours and significant amount of stress worth it.

To my many family members who collectively believed in me, always encouraged my pursuit of education, and instilled a hard work ethic in me. You have always been some of my biggest cheerleaders and I cannot express how grateful I am to have you all in my life. A special thank you to my mom for answering all of my phone calls over the past five years and doing your best to understand the plights of graduate school. I wouldn't have made it through some of the bumps of graduate school without you.

Finally, to the members of my cohort—Sarah, Huan, and Elizabeth—whose friendship and knowledge helped me transition into graduate school and survive the first several years. This journey would have looked a lot different without you all in it. And to all my fellow graduate students, thank you for walking alongside me, mentoring me, encouraging me, and listening. I am so lucky to have found all of you in the chaos of graduate school. I know all of you will continue to do amazing things in your future careers.

Wherever and whenever you joined me on this journey, each and every one of you have shaped my life in a positive way. "Thank you" does not do you all justice.

ABSTRACT

Figurative language appears to be a viable method for examining one's abstract reasoning abilities, as it requires one to make interpretations and inferences based on social context and prior knowledge. Deficits in figurative language have been linked to a variety of clinical conditions and a range of negative psychosocial outcomes. However, in clinical practice there are few measures available to evaluate these language abilities, and the few that do exist are rarely utilized. Previous research indicates that idioms, proverbs, and metaphors are types of figurative language that may prove to be clinically useful in assessing not only abstract reasoning abilities, but also a broader range of cognitive functions. As such, these relationships were examined through two studies. The first study assessed figurative language interpretation in a cognitively-healthy young adult population to describe the range of typical interpretations, and to determine which figurative phrases have the best psychometric utility. Study two utilized the test items identified in study one to examine the relationship between figurative language interpretation and cognitive abilities, including abstract reasoning, cognitive flexibility, working memory, and inhibition.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

Table Page

LIST OF MODELS

Model Page

GLOSSARY OF TERMS

- **Abstract reasoning:** the ability to quickly reason with information to solve novel, unfamiliar problems, independent of any prior knowledge, and the ability to reach logical conclusions in the absence of physical data or concrete phenomena (Marini & Case, 1994; Medin et al., 2002).

- **Cognitive flexibility:** a subdomain of executive function; **t**he ability to appropriately adjust one's behavior according to a changing environment by efficiently disengaging from a previous task, reconfiguring a new response set, and implementing this new response set to the task at hand; most commonly measured with task switching and set shifting tasks (Armbruster et al., 2012; Scott, 1962).

- **Executive functions (EF):** a set of processes aimed at regulating and directing behavior through the management of strategies and plans, the creation of action hierarchies, and the monitoring and adaptation of strategies to different contexts (Delis et al., 2001a).

- **Figurative language:** phrases or statements that convey meaningful information, but are not literally true, and in which context plays an important role in determining the true communicative function of the phrase (Arcara & Bambini, 2016; Merriam-Webster, n.d.).

- **Idioms:** conventional language expressions whose meaning cannot be derived from the literal interpretation of the words, and whose definitions are typically connected to culture, and have a more abstract, and socially-agreed-upon definition (Papagno, 2010).

- **Inhibition:** a subdomain of executive function; the ability to control one's attention, behavior, or thoughts and suppress distracting information and unwanted responses in order to override competing cognitions (Diamond, 2013; Gilmore et al., 2015).

- **Metaphors:** a sentence or phrase that aims to indicate that a property of one item is being conceptually transferred or applied to another item (Iskandar, 2014; Ortony, 1993).

- **Pragmatic language:** the effective use of language to communicate with others in various social contexts (Paltridge, 2021; ASHA, n.d.c).

- **Proverbs:** succinct, concrete phrases that convey a deeper, abstract meaning which express well-known truth based on common sense, social norms, moral concerns, or practical experience (Leyhe et al., 2011; Gibbs & Beitel, 1995; Delis et al., 2001b).

- **Set shifting:** a type of lower-level cognitive flexibility; a type of task that requires individuals to follow one set of rules to complete a task, then shift to using a different set of rules to complete the task (Yerys et al., 2015).

- **Working memory:** a subdomain of executive function; a limited-capacity system that monitors and manipulates information in mind and functions as short-term storage of information (Joormann, 2010; Gilmore et al., 2015; Cohen et al., 1997).

Introduction

It may not take more than a quick walk across our campus to hear a phrase like, "That new Drake album *slaps*[1], it's been *living rent free* in my mind." As just demonstrated, everyday communication not only involves literal language, but also is riddled with figurative (or non-literal) language that can often leave room for interpretation (and, at times, miscommunication). Figurative language is defined as phrases or statements that convey meaningful information, but are not literally true (Merriam-Webster, n.d.). A statement or phrase is considered "figurative" when the gap between the literal meaning and the implied meaning differs, and in which context plays an important role in determining the true communicative function of the phrase (Arcara & Bambini, 2016).

Common examples of figurative language include but are not limited to metaphors, proverbs, idioms, jokes, and irony (Iskandar, 2014). Existing research on idiom, proverb, and metaphor interpretation supports a substantial link between these types of figurative language, various cognitive abilities, and a range of cognitive and psychiatric conditions (Kempler et al., 1988; Papagno et al., 2003; Papagno et al., 2006; Titone et al., 2002; Van Lancker & Kempler, 1987; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Leyhe et al., 2011; Cardoso et al., 2014; Murphy et al., 2013; McDonald et al., 2007). Additionally, idioms, proverbs, and metaphors are believed to be a viable and valuable method of assessing verbal abstraction skills, as they 1) are generally brief, 2) at face-value, they appear to be concrete phrases, and 3) they convey a deeper, abstract meaning which individuals are required to reason through and decode. As such, idioms, proverbs, and metaphors were chosen as the primary constructs of interest in the present study.

[1] Examples of figurative language will be italicized throughout this document.

Proverbs are succinct, concrete phrases that convey a deeper, abstract meaning (Delis et al., 2001b), while metaphors are defined as a sentence or phrase that aims to indicate that a property of one item is being conceptually transferred or applied to another item (Iskandar, 2014). Idioms, on the other hand, are conventional language expressions whose meaning cannot be derived from an analysis of the literal meaning of the individual words within the expression (Papagno, 2010).

Figurative language interpretation, through the use of proverbs, has been utilized for psychological research and clinical purposes in the assessment of abstract reasoning for over a century (Mieder, 1978, Gorham, 1956a, Delis et al., 2001b); however, less information is available on measures of figurative language beyond proverbs (i.e., idioms and metaphors). Existing research on metaphor and idiom interpretation also supports a substantial link between these types of figurative language and abstract reasoning abilities, although there are very few published idiom or metaphor measures and no known clinically-used measures (Kempler et al., 1988; Papagno et al., 2003; Rassiga et al., 2009; Papagno et al., 2006; Papagno & Caporali, 2007; Van Lancker & Kempler, 1987; Kempler et al., 1999; Papagno & Vallar, 2001; Titone et al., 2002; Papagno et al., 1995; Amanzio et al., 2008; Monetta & Pell, 2007; Giora et al., 2012; Iakimova et al., 2006; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007). Further, no known studies have examined whether idioms, proverbs, and metaphors are equally good measures of abstract reasoning, or whether one or two types of figurative language interpretation are superior measures of abstract reasoning.

Beyond abstract reasoning, a cluster of different executive functions (EF; e.g., set shifting, planning, working memory, inhibition) have also been broadly associated with idiom, proverb, and metaphor interpretation (Brune & Bodenstein, 2005; Spoheim et al., 2003; Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008; Monetta & Pell, 2007; Iskandar,

2014). However, no known studies have explicitly examined the strength of the relationships between idiom, proverb, and metaphor interpretation as they relate to different EF abilities.

In summary, idiom, proverb, and metaphor interpretation are known to be related to abstract reasoning abilities. However, to date, proverbs have been given the most research and clinical attention. No clinically-used measures of metaphor or idiom interpretation exist at present, and no study has explicitly examined whether idioms, proverbs, and metaphors have comparable levels of effectiveness at measuring abstract reasoning. Additionally, a cluster of other cognitive abilities are associated with idiom, proverb, and metaphor interpretation, although no known studies have comprehensively compared the strength of the relationships between these three types of figurative language and other cognitive abilities.

Purpose of the Present Study

The present study aimed to fill these gaps in the literature through the conduction of two related studies. Study one aimed to develop a figurative language interpretation task, which included a range of idiom, proverb, and metaphor items which were utilized in study two. The study accomplished these purposes by assessing cognitively-healthy young adults on 60 different figurative language interpretation items. The resulting Figurative Language Interpretation Task was the primary instrument of interest in study two.

Study two aimed to assess how interpretation of idioms, proverbs, and metaphors relate to abstract reasoning, and to examine the strength of the relationship between these types of figurative language interpretation and abstract reasoning. Additionally, this study aimed to explore the relationship between different types of figurative language interpretation and other cognitive measures through the use of pre-existing neuropsychological tests that assess abstract reasoning, inhibition, cognitive flexibility, and working memory.

The results of this exploration were intended to provide additional evidence for the utility of these three types of figurative language interpretation as psychological assessment tools. It was hypothesized that one or two types of figurative language interpretation may be superior measures of abstract reasoning and (or) superior measures of cognitive functioning (as measured by inhibition, intelligence, cognitive flexibility, and working memory tasks).

Literature Review

Background Context of Figurative Language

Pragmatic Language

Figurative language interpretation falls under a larger set of communication skills called "pragmatic language." Pragmatic language refers to the effective use of language to communicate with others in various social contexts (Paltridge, 2021; ASHA, n.d.c). Pragmatic skills include both verbal (e.g., oral speech, written language, prosody, tone, etc.) and nonverbal forms of communication (e.g., eye contact, body language, etc.). Of interest to the present study are verbal pragmatic skills, which typically rely on higher-order linguistic skills, as well as higher-order cognitive abilities (ASHA, n.d.a). Verbal pragmatic language encompasses a variety of linguistic and cognitive abilities. Linguistically, one must understand speech sounds, vocabulary, grammar, prosody, and tone (ASHA, n.d.a). Cognitively, one must be able to take another's perspective, understand and act on social norms, follow social and linguistic rules, decode figurative language, and draw inferences from what is not explicitly stated (ASHA, n.d.a; ASHA, n.d.c). Additionally, language must be interpreted in relation to previous knowledge, the context of the conversation, and within the schema of social cues from the co-communicator (Martin & McDonald, 2003; ASHA, n.d.a). As such, pragmatic language relies on the ability to simultaneously account for both linguistic (e.g., vocabulary, grammar, etc.) and contextual

information (e.g., social, situational, general world knowledge, etc.) (Paltridge, 2021; ASHA, n.d.a). For example, many statements, such as the phrase *"My car broke down again"* may be used for different communicative purposes. On one hand, this phrase may be used for the purpose of complaining about the car (and so we label the phrase a 'complaint'). However, that same phrase — *"My car broke down again"* — may also be used for the purposes of explaining and apologizing for why a person is late (and so we label the utterance an 'explanation'). As demonstrated in this example, the primary purpose of a single utterance can vary significantly based on the social context in which it is used, which is why pragmatic language skills are critical in helping the listener infer the true meaning of the statement.

Clinical Significance of Pragmatic Language

The importance of pragmatic language proficiency for social interaction becomes evident when considering the number of clinical syndromes in which this ability is impaired. Difficulties with pragmatic language are commonly seen in a variety of developmental disorders, including intellectual, developmental, and learning disabilities, autism spectrum disorder, and attention-deficit hyperactivity disorder (APA, 2013; ASHA, n.d.b). Pragmatic deficits are also known to arise in the context of various acquired disorders, including traumatic brain injuries, aphasias, dementias, and right-hemisphere damage (APA, 2013; ASHA, n.d.b; Stemmer, 2008; Bambini, 2010; Bambini & Bara, 2012).

Relatedly, it is well documented that difficulties with pragmatic language skills can result in far-reaching problems, including difficulty participating in social settings (St Clair et al., 2011), developing peer relationships (Whitehouse et al., 2009), achieving academic success (ASHA, n.d.b), and performing successfully on the job (Lewis et al., 2008; Eaves & Ho, 2008). Similarly, pragmatic proficiency has been negatively correlated with social-emotional and

behavioral difficulties, and positively correlated with peer acceptance and the ability to engage in collaborative learning assignments (Gottman et al., 1975; Helland et al., 2014; Kemple et al., 1992; Murphy et al., 2014).

Given the variety of disorders and problems associated with pragmatic impairment, prevalence rates of pragmatic impairment vary considerably depending on the specific population. However, estimates based on a community sample of over 1,300 kindergarteners suggests that pragmatic language impairment occurs in around 7.5% of children (Ketelaars et al., 2009), and another study of eighth graders reported pragmatic language difficulties in 7% to 11% of students (Ellis Weismer et al., 2021). Additionally, prevalence rates may be as much as three times higher for individuals with additional language disorders, and this type of language impairment is more prevalent in boys than girls by a ratio of 5:2 (Ellis Weismer et al., 2021). Taken together, we see that pragmatic language deficits, which impact an estimated 7 to 11% of the population, are associated with a wide range of clinical disorders and with many adverse psychosocial outcomes.

Figurative Language

Description of Figurative Language

Figurative language, a subtype of pragmatic language, is defined as phrases or statements that convey meaningful information, but are not literally true (Merriam-Webster, n.d.). A statement or phrase is considered "figurative" when the gap between the literal meaning and the implied meaning differs, and in which context plays an important role in determining the true communicative function of the phrase (Arcara & Bambini, 2016). These phrases or statements draw upon many of the higher-order cognitive and linguistic skills associated with pragmatic language. Specifically, figurative language requires individuals to integrate a wide range of

contextual and linguistic information in order to reach the intended meaning (Arcara & Bambini, 2016). Because of this, figurative language interpretation may be a useful avenue for assessing higher-order cognitive abilities.

Common examples of figurative language include but are not limited to metaphors, proverbs, idioms, jokes, and irony (Iskandar, 2014). As such, figurative language can come in highly varied forms (Nunberg et al., 1994; Cacciari & Glucksberg, 1995; Gibbs & Colston, 2006). For example, some nonliteral expressions convey plausible actions (e.g., *kick the bucket*, *spill the beans*), while others represent implausible actions (e.g., *wear one's heart on one's sleeve*). Similarly, some expressions are ambiguous and can be interpreted literally as well as figuratively (e.g., *a red eye flight*), whereas others have only a figurative interpretation (e.g., *cost an arm and a leg*). Also, some nonliteral expressions are semantically opaque—difficult to understand based on the literal interpretation of the words (e.g., *kick the bucket*), while others are more semantically transparent—feasibly understood based on the literal interpretation of the words (e.g., *skate on thin ice*; Papagno & Genoni, 2004). In short, figurative language can vary widely in how easily interpreted it is based on its plausibility, literality, and transparency. This range in difficulty also makes figurative language interpretation a suitable candidate for assessing a range of cognitive abilities.

History of Figurative Language and Psychology

Clinical psychology has experienced several waves of exploring figurative language use, initially dating back to the very early days of modern psychology. The first known figurative language interpretation test—a proverb test—was designed to measure intelligence, and was first used in 1906 (Mieder, 1978). From there, additional records have indicated that several updated iterations of proverb tests existed between 1914-1927 (Mieder, 1978). However, the use of

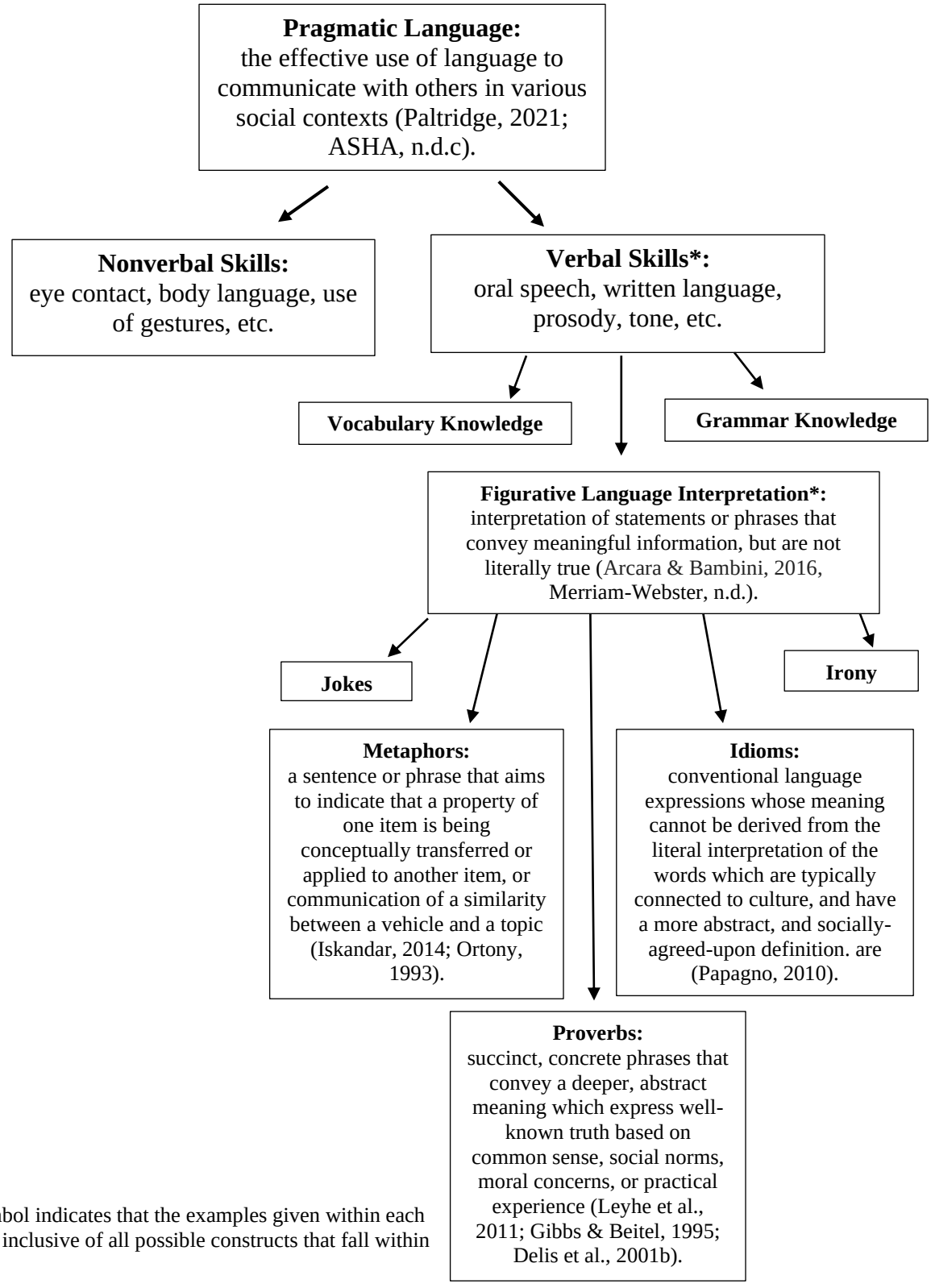
proverbs in clinical practice is most commonly associated with Gorham's (1956a) Proverb Test. Initial research on this measure focused on abstract reasoning abilities and on its capacity to identify individuals with psychotic symptoms (Gorham, 1956a; Gorham, 1956b). Gorham's Proverb Test was later incorporated into clinical practice through its publication as a measure of abstract reasoning within the Delis-Kaplan Executive Functioning System (D-KEFS, Delis et al., 2001a). In summary, we see that figurative language interpretation, primarily through the use of proverbs, has been utilized for psychological research and clinical purposes for over a century, however, very little information is available on other types of figurative language beyond proverbs.

Types of Figurative Language

Figurative language includes but is not limited to metaphors, proverbs, idioms, jokes, and irony, as each of these types of language requires individuals to integrate contextual information and linguistic content in order to reach the intended meaning (Iskandar, 2014; Arcara & Bambini, 2016). Existing research on idiom, proverb, and metaphor interpretation supports a substantial link between these types of figurative language, and various cognitive abilities and psychiatric conditions. As such, idioms, proverbs, and metaphors were chosen as the primary constructs of interest in the present study. Previous literature for each of these subtypes of figurative language is reviewed below. See Model 1 for an overview of the relationship between pragmatic language, figurative language, and idioms, proverbs, and metaphors. Additionally, see Table 1 for examples of idioms, proverbs, and metaphors.

Model 1.

Relationship between Pragmatic Language, Figurative Language, and Idioms, Proverbs, and Metaphors



Note: The * symbol indicates that the examples given within each category are not inclusive of all possible constructs that fall within that category.

Table 1.
Examples of Figurative Language Utilized in the Present Study

Type	Example Phrase	Classification Properties (transparency; plausibility; interpretation)
Idiom	*To play with fire*	Transparent; plausible; ambiguous interpretation
Idiom	*Pass with flying colors*	Opaque; implausible; figurative interpretation
Proverb	*Actions speak louder than words*	Transparent; implausible; figurative only
Proverb	*A watched pot never boils*	Opaque; plausible; ambiguous interpretation
Metaphor	*A judge is a balance*	Transparent; N/A; figurative interpretation
Metaphor	*Hard work is a ladder*	Transparent; N/A; ambiguous interpretation

Proverbs.

Definition. Proverbs are succinct, concrete phrases that convey a deeper, abstract meaning (Delis et al., 2001b). They are known to be simple, familiar sayings which express well-known truth based on common sense, social norms, moral concerns, or practical experience (Leyhe et al., 2011; Gibbs & Beitel, 1995). To convey their message, most—but not all—proverbs make explicit or implicit comparisons between ideas from different, usually unrelated knowledge domains (Gibbs, 1999; Uekermann et al., 2008). For example, the phrase *"every bread has its crust"* has a literal meaning—bread always has crust—but socially, this phrase is used to convey the more abstract meaning that even good things are not without flaw or a negative side. Due to their primary purpose of conveying well-known truths or social beliefs, they are among the most commonly used forms of figurative language.

Ties to Psychosocial and Neurocognitive Functioning. The earliest published studies of proverb interpretation were conducted on patients with schizophrenia (Gorham, 1956b); however, more recent studies have expanded into different neurocognitive and psychiatric patient populations. Initial studies on patients with schizophrenia demonstrated that these individuals

have impaired proverb interpretation (Gorham, 1956b), which sparked significant interest in the study of proverb interpretation as a form of measuring abstract thinking abilities. Many studies repeatedly demonstrated that these individuals selected abstract, but incorrect interpretations for the given proverbs (Thoma et al., 2009; Meadow et al., 1953; Elmore & Gorham, 1957; Haas et al., 2015; Kiang et al., 2007). These deficits in proverb interpretation have been shown in both comprehension and production domains and are thought to be related to the symptoms of executive dysfunction associated with schizophrenia (Haas et al., 2015; Kiang et al., 2007; Thoma et al., 2009).

Since then, several studies have shown that individuals with various neurocognitive disorders also demonstrate impairment in proverb interpretation. Specifically, studies of patients with Alzheimer's disease have found that these individuals tend to produce concrete, rather than abstract interpretations of proverbs (Kempler et al., 1988; Leyhe et al., 2011). Additionally, patients with amnestic mild neurocognitive disorder performed worse than healthy controls on tasks of proverb interpretation, providing concrete and, at times senseless, responses at a higher rate (Leyhe et al., 2011; Cardoso et al., 2014). Parkinson's disease has also been linked to impaired proverb comprehension (Levin et al., 1989). In all of these cases, however, it is worth noting that proverb comprehension deficits also emerge in the course of normal aging, which researchers believe to be related to reduced executive skills (Uekermann et al., 2008).

In addition to these neurocognitive disorders, several different brain lesions and neurologic disorders have also been associated with poor proverb interpretation abilities. For example, in a study examining frontal lesions, Murphy and colleagues (2013) found that medial and left frontal lesions were associated with significantly impaired proverb interpretation while right frontal lesions were associated with proverb interpretation errors that approached statistical

significance. Interestingly, although all groups made interpretation errors, only the left frontal lesion group was associated with higher rates of concrete responses (Murphy et al., 2013). Similarly, patients with frontal lobe epilepsy (FLE) and temporal lobe epilepsy (TLE) have demonstrated impaired performance relative to healthy controls on proverbs tests, with left FLE showing the most inaccuracy in their responses (McDonald et al., 2007). Additionally, the FLE patients demonstrated more impairment in their abstraction abilities than patients with TLE (McDonald et al., 2007). Together, these studies on FLE and left and right frontal lesions suggest that the frontal cortex plays an integral role in abstraction abilities and accurate proverb interpretation. Additionally, coordination and communication between the left and right hemisphere appears to play an important role in proverb interpretation. For example, one study has also found that lack of growth and development in the corpus callosum appears to lead to significant difficulty in proverb interpretation (Rehmel et al., 2016). Taken together, these findings indicate that proverb interpretation, used as a measure of abstract reasoning abilities, may be a valid and clinically useful form of assessing cognitive and psychological deficits.

Metaphors.

***Definition*.** A metaphor is defined as a sentence or phrase that aims to indicate that a property of one item is being conceptually transferred or applied to another item (Iskandar, 2014). Similarly, according to traditional psycholinguistic models (Ortony, 1993), a metaphor is defined as the communication of a similarity between a vehicle and a topic. For example, in a metaphor such as "*love is a battlefield*," "*battlefield"* is the vehicle and "*love"* is the topic. The similarity that the reader infers between these two concepts is the metaphor. A metaphor also can be defined through its function. Metaphors, therefore, function as bridges between two concepts

and create new ideas that transcend what these concepts represent individually (Muran & DiGiuseppe, 1990).

Ties to Psychosocial and Neurocognitive Functioning. Similar to proverb interpretation, metaphor interpretation has also been linked to a wide range of cognitive and psychiatric disorders. Specifically, many individuals with neurocognitive disorders have demonstrated impaired performance on metaphor interpretation tasks. For example, autism spectrum disorder (ASD) has been associated with differences in metaphorical language processing. Specifically, individuals with ASD demonstrate poorer performance on metaphor tasks, display different patterns of brain activation when interpreting metaphors than healthy controls, and may also demonstrate comorbid difficulties with literal language interpretation and theory of mind tasks (Happe, 1993; Gold & Faust, 2010; Giora et al., 2012). Furthermore, individuals with language-based learning disabilities (LDs) tend to have more trouble with metaphorical language than literal language (Jones & Stone, 1989; Lee & Kamhi, 1990; Nippold & Fey, 1983). Interestingly, providing additional context does not appear to improve performance for those with an LD, which is contrary to findings with healthy controls (Lee & Kamhi, 1990). Furthermore, one study has also found Down Syndrome to be associated with impaired metaphorical processing even in the presence of typical performance on general language tasks (Papagno & Vallar, 2001). These studies on ASD, language-based LDs, and Down Syndrome suggest that both intact literal language processing abilities and higher-order cognitive abilities (e.g., contextual processing and perspective taking abilities) are necessary for accurate metaphor interpretation.

Beyond these developmental neurocognitive disorders, several acquired neurocognitive disorders have also been linked to poorer performance on metaphor interpretation tasks. Particularly, individuals with Alzheimer's disease and individuals with Parkinson's disease

demonstrate impairment in metaphor interpretation, likely due to verbal and executive dysfunction commonly seen with these disease processes (Papagno, 2001; Amanzio et al., 2008; Monetta & Pell, 2007). In addition, right-hemisphere brain damage (RBD) has been associated with more incorrect metaphorical interpretations (Winner & Gardner, 1977; Brownell et al., 1990).

Beyond neurocognitive dysfunction, a few psychiatric conditions have also been associated with impaired metaphorical processing. For example, one study found that individuals with depression performed worse on metaphor interpretation tasks than healthy controls, with severity of depressive symptomology and psychomotor retardation being associated with more inaccurate responses (Iakimova et al., 2006). Metaphor interpretation is also impaired in patients with schizophrenia, with patients being more likely to provide literal or concrete responses (Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007). Further, the severity of formal thought disorder and the presence of an acute psychotic episode have been associated with higher degrees of impairment in metaphor interpretation. Taken together, similar to proverb interpretation, metaphor interpretation appears to measure abstract reasoning abilities, and may be a valid and clinically useful form of assessing cognitive and psychological deficits.

Idioms.

***Definition*.** Idioms are conventional language expressions whose meaning cannot be derived from an analysis of the literal meaning of the individual words within the expression (Papagno, 2010). These statements are typically deeply connected to culture, and have a more abstract, and socially-agreed-upon definition. For example, the phrase *"it's raining cats and dogs"* could have a literal (although implausible) interpretation—that cats and dogs are literally

falling from the sky. However, the socially-known interpretation of this phrase is that the rain is very heavy. Idiomatic expressions are among some of the most commonly used forms of figurative language, as they function to convey general information using socially or culturally known definitions (Papagno, 2010; Jackendoff, 1995). Because of this, some idioms are thought to be stored in semantic memory together with word meanings and concepts, as well as other common word strings, such as book titles, lyrics, and clichés (Jackendoff, 1995).

Ties to Psychosocial and Neurocognitive Functioning. As with proverb and metaphor interpretation, idiom interpretation has also been linked to a wide range of psychiatric and cognitive disorders. Specifically, several types of acquired neurocognitive dysfunction have been associated with impaired abstraction abilities on idiom interpretation tasks. For example, individuals with Alzheimer's disease demonstrate impairment in idiom interpretation, wherein concrete and literal responses are more likely to be produced than abstract responses (Kempler et al., 1988; Papagno et al., 2003; Rassiga et al., 2009). In addition, individuals with right-hemisphere (RBD) and left-hemisphere brain damage (LBD) can also demonstrate impaired idiom comprehension, although the pattern of impairment varies based on the location of lesion (i.e., RBD patients have normal literal interpretations and impaired figurative interpretations; Van Lancker & Kempler, 1987; Kempler et al., 1999).

Additionally, disorders that affect language processing abilities have also been associated with impaired idiom comprehension. Specifically, patients with aphasia (both fluent and non-fluent) performed worse than healthy controls on tasks of idiomatic interpretation, selecting inaccurate and literal responses at a higher rate (Papagno et al., 2006; Papagno & Caporali, 2007). Theoretically, it is believed that since language resources are damaged in aphasics, a greater involvement of executive control is required in linguistic tasks, thus depleting the

attentional pool, and preventing the appropriate suppression/inhibition of the literal meaning (Papagno et al., 2006; Papagno & Caporali, 2007). One study also found a connection between Down Syndrome and impaired idiom interpretation even in the presence of typical performance on general language tasks (Papagno & Vallar, 2001). These studies suggest that both intact literal language processing abilities as well as higher-order cognitive abilities (e.g., inhibition and selection processes) are integral components of idiom comprehension.

Furthermore, idiomatic language processing has also been shown to be impaired in patients with Schizophrenia, with patients demonstrating difficulty in producing abstract and nonliteral responses (Titone et al., 2002; Papagno et al., 1995). These difficulties are thought to be associated with lower abstract thinking abilities and with difficulty in inhibiting contextually inappropriate responses (Titone et al., 2002; Papagno et al., 1995). In summary, similar to proverb and metaphor interpretation, idiom interpretation appears to measure abstract reasoning abilities, and may be a valid and clinically useful form of assessing cognitive and psychological deficits.

Cognitive Models of Figurative Language Processing

There are two main hypotheses used to describe how figurative language is cognitively processed. The first, more traditional model is referred to as the sequential approach or the indirect approach, which suggests that figurative language processing occurs in a series of steps. In the first step, all phrases are processed literally, and in the second step, the meaning of the statement is judged for accuracy (Searle, 1979). If the literal meaning is found to be inaccurate based on context, it is subsequently suppressed and sent to a non-literal processing center (Papagno, 2010). In the third step, a higher-level process occurs in order to decode and infer the nonliteral meaning of the phrase (Searle, 1979). This model posits that understanding

figurative language will always be more effortful than understanding literal language, and that literal interpretations will always take priority over nonliteral interpretations (Iskandar, 2014; Papagno, 2010). Further, to reach the accurate nonliteral conclusion through the indirect or sequential approach, it can be inferred that an inhibitory process likely occurs to prevent acting on the first interpretation and to provide a mental working space for further decoding and processing.

The second model is referred to as the direct approach or the parallel processing approach (Honeck & Hoffman, 1980). The direct model posits that literal processing and nonliteral processing occur through parallel mental search processes rather than through successive mental search processes. In this model, to reach accurate nonliteral interpretations divided attention and working memory would likely play an important role, as these abilities allow an individual to process information through separate channels or systems (Mordkoff & Yantis, 1991).

Some researchers have explored these theories by attempting to distinguish between automatic and effortful processing of figurative language, and most support has been in favor of the direct/parallel processing approach. One study found that the degree of processing (i.e., automaticity or effortfulness) of metaphorical language is largely dependent on context (e.g., the phrase "*let the cat out of the bag"* is automatically understood as figurative when preceded by a phrase like "*the surprise was spoiled by Sarah, she…;"* Ortony, 1978). Other studies have found differences in response times between different types of figurative language and between literal and nonliteral language (Johnson, 1996; Glucksberg et al., 1982). For example, Johnson (1996) found that participants responded quicker to certain metaphors ("*cigarettes are timebombs"*) as compared to their simile counterparts ("*cigarettes are like timebombs"*). These findings are important, as it suggests that not all figurative language is processed through the sequential or

indirect approach, as one would expect all types of figurative language to elicit the similar response times in that model.

However, a more realistic explanation may be that different types of figurative language are processed differently depending on the characteristics of the particular phrase or statement. For example, some researchers theorize that when the phrase is familiar or simple, both its literal and figurative meanings are equally activated (Graded-Salience Hypothesis; Giora, 1997). On the other hand, when it is novel or more complex, the literal meaning is more likely to be activated than the figurative meaning. In support of this idea, studies have shown the figurative meaning of highly familiar metaphors is processed before the literal meaning (e.g., *"She is a warm person"* was more quickly understood as a personality trait than an actual temperature; Blasko & Connine, 1993; Giora et al., 2012). This may suggest that some figurative phrases (familiar or overlearned phrases) are encoded into long-term memory as entire semantic units (i.e., vocabulary terms), and no longer require the mental search to find a similarity between the two concepts (Iskandar, 2014). Alternatively, the mental search may occur as an automatic process, which also would bypass the mental search step. On the other hand, novel or complex phrases likely still require the use of higher-order cognitive processes. In conclusion, there appear to be various factors that may influence the manner in which figurative language is cognitively processed, and characteristics such as context, degree of complexity, and degree of familiarity likely mediate which type of processing occurs.

Neuroimaging Correlates of Figurative Language Processing

Early studies on figurative language comprehension emphasized the role of the right hemisphere (McDonald, 1999; Brownell et al., 1990; Pobric et al., 2008). More recent studies have expanded upon this finding, and it is now believed that rather than strict lateralization of

functions among brain regions, a widely distributed network involving both hemispheres contributes to figurative language processing (Lee & Dapretto, 2006; Rapp et al., 2004; Rapp et al., 2007; Stringaris et al., 2007; Mashal & Faust, 2010). This is likely the case as figurative language interpretation appears to follow the same procedure as literal language interpretation, plus a selection and monitoring process between alternative meanings (Balconi, 2010). The processing of literal language is known to primarily activate the left hemisphere and bilateral temporal lobes as well as Broca's area for speech production and Wernicke's area for language comprehension. EF-related abilities, including the aforementioned communication selection and monitoring processes, are generally believed to be housed within the frontal lobe, and more specifically within the prefrontal regions (Bottini et al., 1995; Balconi, 2010). Based on this information, one would expect to see diffuse patterns of activation in the right hemisphere, frontal regions, and throughout the temporal lobes.

Additionally, previous research supports an increased pattern of brain activation for figurative language interpretation. Specifically, studies assessing event-related brain potentials (ERPs) have demonstrated that more processing (i.e., more brain activation) was required to understand nonliteral sentences than literal ones (Pynte et al., 1996; Coulson & Van Petten, 2002; Sotillo et al., 2004). In addition, there appears to be a pattern of higher left hemisphere activation with nonliteral sentence interpretation when compared to literal sentence interpretation (Yi et al., 2017).

Furthermore, there are some specific brain regions that appear to be activated for figurative language interpretation. For example, findings from functional magnetic resonance imagining (fMRI) studies have demonstrated that proverb interpretation (Bohrn et al., 2012; Kaiser et al., 2013), metaphor analysis (Bottini et al., 1994), and idiom processing (Lauro et al.,

2008; Zempleni et al., 2007; Hillert & Buracas, 2009) lead to increased activation in the prefrontal region. Additionally, results from several repetitive transcranial magnetic stimulation (rTMS) experiments indicate that the prefrontal region plays an important role in idiomatic processing (Fogliata et al., 2007; Rizzo et al., 2007; Oliveri et al., 2004). Additionally, fMRI studies have demonstrated that proverb interpretation (Bohrn et al., 2012; Kaiser et al., 2013), metaphor interpretation (Shibata et al., 2007; Eviatar & Just, 2006), and idiom interpretation (Lauro et al., 2008; Zempleni et al., 2007; Hillert & Buracas, 2009) lead to increased activation in areas of the left temporal lobe. Support for right temporal lobe activation for metaphor comprehension (Bottini et al., 1994) and idiom processing (Lauro et al., 2008; Zempleni et al., 2007; Hillert & Buracas, 2009) has also come from fMRI studies. Furthermore, both Broca's area and Wernicke's area have been found to be activated when an individual is tasked with interpreting figurative language (Lee & Dapretto, 2006; Rapp et al., 2004; Rapp et al., 2007; Stringaris et al., 2007; Mashal & Faust, 2010). Taken together, these studies appear to support the idea that a diffuse network of language-related (left hemisphere, bilateral temporal lobes) and executive function-related brain regions (the prefrontal areas) are involved in the processing of figurative language.

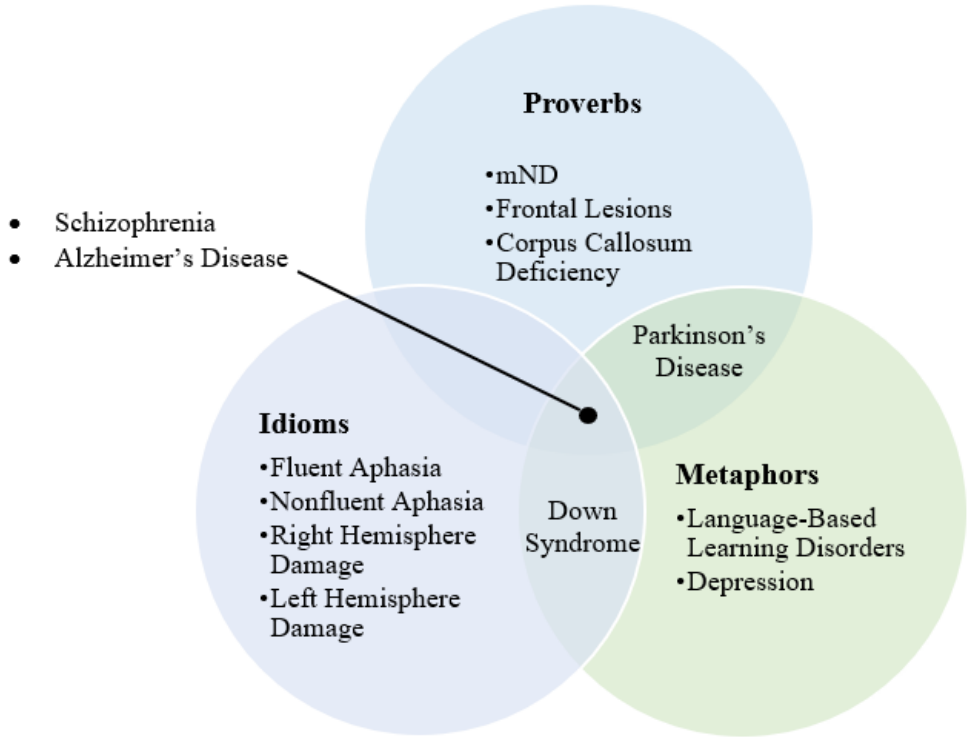
Comparison of Types of Figurative Language

Differences in Definition. Significant overlap is present among the various types of figurative language; however, it is also worth noting the differences between idioms, proverbs, and metaphors. Metaphors differ from proverbs and idioms, as metaphors do not necessarily have a standardized meaning among speakers, but instead can convey more than one meaning and are somewhat more subjective, or based on individual interpretation (Canal et al., 2010). Idioms, on the other hand, do have a unique, standardized meaning among speakers of the

language that can also be customized or uniquely implemented based upon context, but the standard meaning does not change (Fellbaum, 2007). Proverbs also function differently, as they are generally true statements—both literally and figuratively—that typically comment on culturally or community-wide shared situations or experiences (Canal et al., 2010).

Differences in Clinical Syndromes. In addition to differences in definition, these three types of figurative language are also associated with different arrays of clinical conditions. These variations may provide supportive evidence for the conclusion that idioms, proverbs, and metaphors involve different types or levels of cognitive processing. However, no studies to date have comprehensively compared each of these types of figurative language as they relate to clinical syndromes or cognitive functions, so it may also be possible that these types of figurative language are more closely related than current evidence suggests. For example, some clinical conditions, such as Alzheimer's disease and schizophrenia, are associated with proverbs, metaphors, and idioms. On the other hand, Parkinson's disease has only been associated with metaphors and proverbs, while Down syndrome has only been associated with metaphors and idioms. There are also a number of conditions that, to date, have only been associated with one form of figurative language interpretation. A summary of this data can be found in Model 2.

Model 2.
Comparison of Clinical Conditions associated with Types of Figurative Language



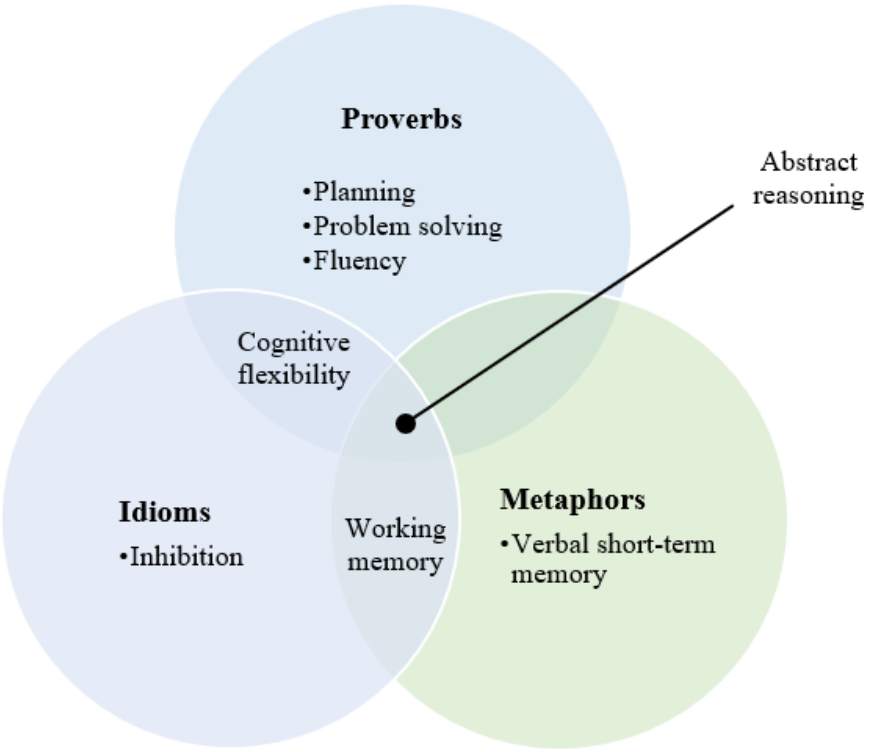
Differences in Cognitive Abilities. There is also substantial variation in previous literature regarding what cognitive abilities are associated with idioms, proverbs, and metaphors. As described above, abstract reasoning abilities are associated with impaired idiom, proverb, and metaphor interpretation. Beyond abstract reasoning, the most common set of cognitive abilities associated with figurative language interpretation are a cluster of different executive functions. Executive functions are typically defined as a set of processes aimed at regulating and directing behavior through the management of strategies and plans, the creation of action hierarchies, and the monitoring and adaptation of strategies to different contexts (Delis et al, 2001b). Executive functions are typically activated in one or more of these situations: (1) when the level of complexity of a task requires more than automatic processing, (2) when old information must be processed in a novel way, or (3) when the incoming information to be processed is novel (Iskandar, 2014). Theoretically, it is believed that figurative language interpretation requires

effortful, conscious effort (Balconi, 2010, Stuss & Alexander, 2000), as less familiar or typical forms of language are associated with higher levels of executive control (Balconi, 2010; Iskandar, 2014). However, there appears to be much variation in the relationships between different executive functions and different types of figurative language.

Specifically, proverb interpretation has been shown to correlate with the EF abilities of planning, problem solving, fluency, and cognitive flexibility (Brune & Bodenstein, 2005; Spoheim et al., 2003). On the other hand, poor idiom comprehension has been correlated with deficits in inhibition, cognitive flexibility, and working memory (Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008). Additionally, impaired verbal working memory and short-term memory have been associated with performance on metaphor comprehension and production tasks (Monetta & Pell, 2007; Iskandar, 2014).

In summary, EF abilities appear to have direct associations with the interpretation of figurative language, as a wide range of EF abilities (e.g., cognitive flexibility, working memory, inhibition, etc.) can be associated with impaired performance on measures of figurative language. However, findings between different types of figurative language (i.e., idioms, proverbs, and metaphors) appear to be inconsistent and variable as to which cognitive abilities they are associated with. Additionally, no known studies have comprehensively examined the relationships (and the strength of the relationships) between cognitive abilities and proverb, metaphor, and idiom interpretation (i.e., studies to date have typically chosen only one form of figurative language and one, two, or three cognitive abilities). A summary of this information can be found in Model 3.

Model 3.
Summary of Cognitive Abilities known to be associated with Figurative Language Interpretation Deficits



Figurative Language in Clinical Practice

Assessment. Figurative language tests are believed to assess an individual's verbal abstract reasoning abilities (Elmore & Gorham, 1957; Delis et al., 2001b; Iskandar, 2014). Idioms, proverbs, and metaphors may represent a viable and valuable method of assessing verbal abstraction skills, as they 1) are generally brief, 2) at face-value, they appear to be concrete phrases, and 3) they convey a deeper, abstract meaning which individuals are required to reason through and decode.

At present, only one form of figurative language interpretation has been created for adult clinical practice—the Proverbs subtest of the Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001a), which was based on Gorham's (1956a) Proverb Test. Although proverb interpretation has been associated with a variety of psychiatric and cognitive disorders,

historically, these tasks have primarily been used as a diagnostic tool to aid in the diagnosis of schizophrenia, and later to differentiate individuals with dementia (Rapp & Wild, 2011). Other forms of figurative language beyond proverb interpretation, such as metaphorical and idiomatic interpretation, have not been included in any clinically used English-language tests for adults (Iskandar, 2014). However, research demonstrates that other forms of figurative language interpretation (i.e., metaphors and idioms) may be uniquely affected by different lesions and clinical conditions. For this reason, metaphor and idiom interpretation may also be suitable methods for assessing verbal abstract reasoning abilities and other cognitive functions (Murphy et al., 2013). Overall, figurative language may be a valuable part of a cognitive profile, although very little attention is given to this domain of functioning at present.

Therapy. Regarding therapeutic practice, studies have found that the use of figurative language, such as metaphors, can enhance understanding between client and therapist, as this type of language allows two different individuals to make common comparisons based on shared experiences or knowledge (Robert & Kelly, 2010; Muran & DiGiuseppe, 1990). Specifically, clients may be encouraged to use metaphors to help them express and structure their experience by connecting it to a similar concept through their own unique perception of the world. Through this, clients do not have to delve into details of their own issues and concerns in the literal sense but can still understand and confront difficult issues in their lives. Therapists may also use metaphors as an indirect method of communication in order to offer their clients ideas, solutions to consider, reframing, or other inputs which may or may not be immediately acceptable to a client (Barker, 1992).

For example, many third-wave cognitive behavioral therapy (CBT) approaches, such as Acceptance and Commitment Therapy (ACT) and other mindfulness-based techniques, employ

the use of metaphors as psychoeducation for clients (Hayes et al., 2016). For example, a therapist may relate negative thoughts to "a hungry tiger," in which the more you feed (i.e., pay attention to) the thoughts, the larger they grow, and subsequently the more you have to feed (i.e., pay attention to) these thoughts, and so on in a self-reinforcing cycle. As another example, mindfulness of thoughts may be related to "leaves on a stream," where the client's thoughts are presented as leaves which the client should allow to flow down a stream (i.e., away from them freely, without continuing to think over them). In summary, we see that, in addition to its utility as a psychological assessment measure, figurative language interpretation plays an important role in the therapeutic setting.

Present Study

Overall, previous research has demonstrated that various forms of figurative language can be uniquely affected by different lesions, clinical syndromes, and cognitive abilities (e.g., Kempler et al., 1988; Papagno et al., 2003; Papagno et al., 2006; Titone et al., 2002; Van Lancker & Kempler, 1987; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Leyhe et al., 2011; Cardoso et al., 2014; Murphy et al., 2013; McDonald et al., 2007). Despite this, figurative language assessments are 1) largely unavailable for use with adult populations in psychological treatment or assessment settings, and 2) what is currently available is not popularly incorporated into clinical practice (Arcara & Bambini, 2016). The most widely available assessment of figurative language interpretation is the D-KEFS Proverb Test, which is primarily used as a measure of abstract reasoning (Delis et al., 2001a). However, other types of figurative language interpretation (i.e., metaphors and idioms) are not incorporated into clinical practice, although current evidence suggests that impaired metaphor and idiom interpretation is associated with

deficits in abstract reasoning (Kempler et al., 1988; Papagno et al., 2003; Rassiga et al., 2009; Papagno et al., 2006; Papagno & Caporali, 2007; Van Lancker & Kempler, 1987; Kempler et al., 1999; Papagno & Vallar, 2001; Titone et al., 2002; Papagno et al., 1995; Amanzio et al., 2008; Monetta & Pell, 2007; Winner & Gardner, 1977; Brownell et al., 1990; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Jones & Stone, 1989; Lee & Kamhi, 1990; Nippold & Fey, 1983; Iakimova et al., 2006; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007).

Beyond abstract reasoning, a wide range of EF abilities (e.g., set shifting, planning, working memory, inhibition, etc.) has been associated with impaired performance on measures of idiom, proverb, and metaphor interpretation (Brune & Bodenstein, 2005; Spoheim et al., 2003; Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008; Monetta & Pell, 2007; Iskandar, 2014). However, current research on the relationships between different types of figurative language—idioms, proverbs, and metaphors—and EF abilities is largely inconsistent.

In summary, idioms, proverbs, and metaphors are each assumed to measure abstract reasoning. However, no known studies have examined whether idioms, proverbs, and metaphors are equally good measures of abstract reasoning, or whether one or two types of figurative language interpretation are superior measures of abstract reasoning. Furthermore, there is mixed and inconsistent evidence regarding how different types of figurative language relate to other cognitive abilities beyond abstract reasoning, such as verbal intelligence, inhibition, cognitive flexibility, and working memory.

The present study aimed to help fill these gaps in the literature through two related studies. Study one aimed to develop a figurative language interpretation task, which included a range of idiom, proverb, and metaphor items. The study accomplished these purposes by

assessing cognitively-healthy young adults on 60 different figurative language interpretation items. Participants' responses to each item were scored based on both accuracy and level of abstraction by two independent raters. The Figurative Language Interpretation Task has a total score and three subscale scores—one for idioms, one for proverbs, and one for metaphors. The resulting Figurative Language Interpretation Task was the primary instrument of interest in study two. The specific aims of study one were to:

1) Ensure inter-rater reliability among co-raters on the Figurative Language Interpretation Task items.
2) Reduce the number of items of the Figurative Language Interpretation Task by removing items that fail to elicit an adequate range of responses or do not demonstrate adequate association with other items on the measure.
3) Obtain acceptable alpha levels for each subscale.

Study two aimed to assess how interpretation of idioms, proverbs and metaphors relate to abstract reasoning, and to examine the strength of the relationship between these three types of figurative language interpretation and abstract reasoning. Additionally, this study aimed to explore the relationship between different types of figurative language interpretation and other cognitive measures through the use of pre-existing neuropsychological tests that assess general intelligence, inhibition, cognitive flexibility, and working memory. Figurative language interpretation was assessed using the Figurative Language Interpretation Task created in study one, which includes a total score, an idiom score, a proverb score, and a metaphor score. The results of this exploration can be utilized to provide additional evidence for or against the utility of figurative language interpretation as a psychological assessment tool. It is believed that one or two types of figurative language interpretation may be superior measures of abstract reasoning

and (or) superior measures of cognitive functioning (as measured by inhibition, intelligence, cognitive flexibility, and working memory tasks), although the mixed and inconsistent literature limits the ability to make strong directional hypotheses at present. As such, there were two aims of study two:

1) Examine the final factor structure of the measure.
2) Assess alpha levels for any identified scales.

Additionally, study two had four experimental hypotheses:

1) As current evidence suggests that figurative language interpretation is linked to abstract reasoning (Elmore & Gorham, 1957; Delis et al., 2001b; Iskandar, 2014), the Figurative Language Interpretation Task scales will collectively demonstrate a positive association with other measures of abstract reasoning.
2) Performances on the idiom, proverb, and metaphor interpretation subscales will each significantly contribute to predicting abstract reasoning performance.
3) The total score, idiom score, proverb score, and metaphor score will each demonstrate a positive association with a) intelligence scores, b) inhibition scores, c) working memory scores, and d) cognitive flexibility scores, as each of these cognitive abilities have been associated with types of figurative language interpretation in previous studies (Brune & Bodenstein, 2005; Spoheim et al., 2003; Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008; Monetta & Pell, 2007; Iskandar, 2014).
4) Male participants and female participants will vary in their figurative language interpretation abilities, with female participants scoring higher on average, as previous research has indicated that the ratio of language impairment between females and males is 2:5 (Ellis Weismer et al., 2021).

Study One

Method

Participants

Based on a power analysis conducted with GPower, study one required 64 participants to achieve 80% power. Accordingly, a total of 64 students from a midsized university in the Midwestern U.S. participated in this study and were compensated with either required research credits for a course or extra credit. Participants were recruited through the SONA Experiment Management System, an online participant management software system, and through advertisements distributed across the university's campus. Individuals were eligible to participate if they were fluent American-English speakers between the ages of 18 and 89 years old. However, participants with a history of severe psychiatric disorder (i.e., Schizophrenia, Bipolar Disorder), neurodevelopmental disorder (e.g., developmental disorder, learning disorder), or brain injury known to impact their cognition were excluded. The total sample (N = 64) was composed of 21 male participants (33%) and 43 female participants (67%). The participants' average age was 24.48 (SD = 8.06) years. There were 44 White/non-Hispanic (68%), 7 African-American/Black (11%), 6 Hispanic/Latino (9%), 4 Asian/Pacific Islander (7%), 1 American Indian/Native Alaskan/Aleutian/Eskimo (2%), and 2 bi-racial (3%) participants. All participants were fluent in American English, however, 56 participants (88%) indicated that English was their primary or native language, while 8 participants (12%) indicated that English was not their primary or native language. Fifty-one participants (80%) reported being fluent in only one language, 11 participants (17%) reported being fluent in two languages, and 2 participants (3%) reported being fluent in three or more languages.

Procedures

Ethical approval was granted by the Wichita State University Institutional Review Board, and informed consent was obtained from all participants before testing. Each study session consisted of one participant with the examiner. Each participant attended the study appointment via the videoconferencing platform Zoom. First, participants were consented into the study, and then the tasks were presented in the following order: the SHIPLEY-1 Verbal Reasoning subtest, the Figurative Language Interpretation Task, and the demographics form. For each task, the examiner read aloud and presented written instructions on a PowerPoint slide that was screenshared with the participant. For the SHIPLEY-1, the Figurative Language Interpretation Task pilot items, and the demographics form, the participants were given as much time as needed to read each item, decide on an answer, and then verbally tell the examiner their answer. The examiner recorded each answer on a data collection form. The study took approximately 40 minutes to complete. After completion of the study, participants were compensated with research participation credits to use for class credit.

Materials

Demographics Form. The demographics form was a self-report questionnaire made by the author, which collects information on age, sex, race/ethnicity, educational level, total years of education, native language, and number of languages the participant is fluent in. The demographics form was created in accordance with the American Psychology Association's guidelines for assessing multicultural demographic information (American Psychological Association [APA], 2017; APA, 2019).

SHIPLEY Institute of Living Scales, Verbal Reasoning Subtest. The SHIPLEY Institute of Living Scales (SHIPLEY-1; Shipley, 1946), Verbal Reasoning Subtest is a 40-item

IQ test that is designed to assess verbal reasoning abilities. Each item prompts the participant to read a target word and then choose from four options which word is closest in meaning to the target word. There is one correct answer, and the other three options are foils. The SHIPLEY-1 has demonstrated good test-retest reliability after a two-month interval (r = .72) and a four-month interval (r = .79) (Goodman et al., 1974). Evidence for convergent validity has been demonstrated through correlations with other measures of intelligence, such as the Weschler Adult Intelligence Scale (r = .81), the Weschler Adult Intelligence Scale – Revised (r = .86), and the Wide Range Vocabulary Test (r = .73; Zachary et al., 1985; Martin et al., 1979; Winkler, 1980; Weiss & Schell, 1991).

Figurative Language Interpretation Task Pilot Stimuli. The Figurative Language Interpretation Task pilot stimuli developed for this study included a total of 60 items—15 familiar (American English) idioms, and 15 unfamiliar (Australian English) idioms ,15 proverbs, and 15 metaphors. The 60-item Figurative Language Interpretation Task pilot stimuli can be found in Appendix A.

Selection of Items for the Figurative Language Interpretation Task. The proverbs were chosen from two studies on proverb understanding (Benjafield et al., 1993; Billow, 1975), which drew many of their proverbs from pre-existing proverb measures used in psychological testing (e.g., Root, 1921; Watts, 1950). Benjafield et al.'s (1993) study asked participants to rate each proverb on a variety of characteristics, one of which was "concreteness," defined as the degree to which the item refers to objects, persons, places, or things that can be perceived through the senses as compared to abstract concepts that cannot be directly experienced. Concreteness was rated on a 7-point Likert scale. For the present study, items were chosen with a moderate level of concreteness with scores ranging from 3.03 to 5.53 (M = 4.38). To aid in the creation of scoring

criteria, standardized definitions for each of these proverbs were used from the Oxford Dictionary of Proverbs—Sixth Edition (Speake, 2015).

The metaphors were chosen from two studies on metaphor understanding (Katz et al., 1988; Campbell & Raney, 2016). Campbell and Raney's (2016) study was a 25-year replication of Katz et al.'s (1988) design, and as such the 2016 data were utilized to choose the metaphors for this study. Campbell and Raney (2016) asked participants to rate each metaphor on a variety of characteristics, two of which were "ease of interpretation," defined as how easily one can interpret the meaning of the phrase, and "metaphor goodness," defined as how good or apt the metaphor is perceived to be. Both ease of interpretation and metaphor goodness were rated on a 7-point Likert scale. For the present study, items were chosen with a moderate level of ease of interpretation (range: 4 to 6, $M = 5.54$) and a moderate level of metaphor goodness (range = 4 to 6, $M = 4.85$). To aid in the creation of scoring criteria, definitions for each of these metaphors were initially modeled from the responses elicited in Iskandar and Baird's (2013) Metaphor Interpretation Task.

The familiar (American English) idioms were chosen from previous studies on idiom understanding (Titone & Connine, 1994; Libben & Titone, 2008; Tanner & Bulkes, 2017), and example sentences were drawn from existing dictionaries (Kelly, 2007; Ayto & Oxford University Press, 2020). In each of the studies on idiom understanding, participants were asked to rate each idiom on a variety of characteristics, one of which was "familiarity," defined as how often the expression is used or heard. For the present study, items were chosen with a moderate level of familiarity with scores ranging from 2.6 to 6.4 ($M = 5.29$). To aid in the creation of scoring criteria, definitions for each of the idioms were used from the Cambridge Idioms

Dictionary—Second Edition (Kelly, 2007) and the Oxford Dictionary of Idioms—Fourth Edition (Ayto & Oxford University Press, 2020).

Idioms are posited to be stored in semantic memory (memory for vocabulary words), and generally have a socially-agreed-upon definition (Jackendoff, 1995). As such, it is possible that idioms may not tap abstract reasoning abilities as well as other forms of figurative language, such as proverbs or metaphors. Instead, well-known idioms may function more like a test of vocabulary knowledge. For this reason, unfamiliar idioms (i.e., Australian English idioms that are less likely to be well-known by American participants) were chosen for this study as an additional option in the case that familiar idioms fail to demonstrate adequate associations with the other constructs of interest. Unfortunately, similar data as utilized in the selection of the other proverbs, metaphors, and familiar idioms was not available for the unfamiliar idioms. As a result, the unfamiliar English idioms were chosen from an online dictionary of Australian slang (Aussie Slang Dictionary, 2022) which provided sample statements as well as general definitions of each phrase, which were utilized in the creation of scoring criteria for the measure.

Subscales and Scoring System for the Figurative Language Interpretation Task. These different types of figurative language—idioms, proverbs, and metaphors—were treated as separate constructs and scales within the Figurative Language Interpretation Task based on their unique definitions and distinct functionality in the English language (i.e., these subtypes of figurative language are classified as separate constructs in the English language; Canal et al., 2010; Fellbaum, 2007). Subsequently, one subscale score was computed for each of the figurative language constructs included in the pilot items – a familiar idiom subscale, an unfamiliar idiom subscale, a proverb subscale, and a metaphor subscale. Additionally, a Total

Achievement Score was calculated by summing the participant's scores on all 60 items and was used to represent an individual's ability to interpret figurative language as a singular construct.

For the administration of this measure, each of the 60 phrases and statements were simultaneously read aloud and visually presented to the participant. After each item was presented, the participant was asked to provide a verbal response of what they believed each phrase or statement meant. The instructions were as follows: "I am going to read some statements and phrases. After each one, tell me what the statement or phrase means. The statements and phrases will also be presented here for you to read. The phrase or statement will be printed in bold, underlined font."

The scoring system for this measure was informed by several pre-existing clinically and experimentally used measures (Albert et al., 1990; Delis et al., 2001a; Wechsler, 1997; Iskandar & Baird, 2013). Based on these sources, verbal responses can vary according to the degree of accuracy and the degree of abstraction, so a scoring system was developed to examine the degree of abstraction and the degree of accuracy of each response. Accuracy is rated according to how well the meaning of the response "fits" the meaning of the statement regardless of the level of abstraction of the response. Accuracy scores can be fully accurate (2 points), partially accurate (1 point) or incorrect (0 points). Abstraction is rated according to the level of generalization (i.e., applies to more situations, people, or concepts than are reflected in the literal meaning of the original statement) represented by the response without regard to its accuracy. Abstract responses are awarded 2 points, while concrete/literal responses are awarded 0 points. These two dimensions are scored independently of each other, and the two scores are combined to form a total score for each response, with a maximum total score of 4 points. Consequently, responses

fall into five general types, as seen in Table 2. Specific scoring criteria can be found in Appendix B.

Table 2.
Description of Scores Possible for the Figurative Language Interpretation Task

Description	Accuracy Score (max: 2)	Abstraction Score (max: 2)	Total Score (max: 4)
Accurate & Abstract	2	2	4
Partial & Abstract	1	2	3
Accurate & Concrete	2	0	2
Partial & Concrete	1	0	1
Incorrect/Concrete	0	0	0

The total score for every item was summed into a single Total Achievement Score for the Figurative Language Interpretation Task pilot stimuli. In addition to the Total Achievement Score, a separate subscale score was computed for each of the individual four subscales using the sum of its respective items' total scores (e.g., all 15 metaphor scores were summed into the Metaphor Subscale score).

In addition to the primary researcher, a graduate research assistant was trained to score the items using sample responses and standardized scoring criteria. The training process involved scoring three protocols together and discussing the reasoning and decision-making process behind each score. After completing this, the primary researcher and the graduate research assistant scored three additional protocols separately, compared their separate scores, and discussed any relevant differences to ensure that scoring aligned and was based on similar reasoning and decision-making processes. At this point, each researcher scored the remaining protocols independently. A random sampling of 20 protocols from the present study was scored by this second rater to ensure inter-rater reliability for all response types.

Data Analysis

Analytic Strategy. Following data cleaning, demographic frequencies were run on answers to the demographic form, as well as scores on the SHIPLEY-1 to examine the sample's characteristics. Additionally, univariate comparisons and independent-samples t-tests using relevant demographic variables and the main scores of interest were conducted to examine potential differences and associations that may influence subsequent analyses and interpretation of the data. Demographic information, the SHIPLEY-1 score, the Familiar Idiom Subscale, Unfamiliar Idiom Subscale, Proverb Subscale, Metaphor Subscale, and Total Achievement Score were included in these analyses.

To address aim (1), ensuring interrater reliability, a random sampling of 20 participant protocols was scored by a second rater and an intraclass correlation was conducted between the two raters' scores. According to Koo and Li, (2016), "values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability."

To address aim (2), reducing the number of items, an item discrimination analysis was computed by correlating each item's average score with the average score obtained on the item's respective subscale and with the Figurative Language Total Score. Previous research indicates that Pearson's *r* correlations of 0.2 or higher are considered acceptable items to maintain in the measure (Schweizer & DiStefano, 2016, pg. 98). In addition, an item difficulty index was calculated (see Schweizer & DiStefano., 2016, pg. 97 for a description) for each item by calculating the average score across all participants for each item. The average score obtained on each item represents the item's "difficulty," with higher scores representing "easier" items and lower scores representing "more difficult" items. The items are then arranged in descending

order of difficulty, and items with different levels of difficulty can be identified and selected for further analysis or further inclusion in the final measure.

To address aim (3), Cronbach's alpha was calculated for each subscale (idiom, proverb, metaphor, and Total Achievement), any potential items that drop the alpha level below the adequate threshold were removed. Cronbach's alpha of 0.70 or higher is considered adequate internal consistency for exploratory research (Hair et al., 2014; Nunnally, 1978). This technique was applied concurrently with the item reduction techniques to produce a measure that is both succinct and psychometrically sound.

Results

Data Cleaning

After data collection was completed, data cleaning procedures were undertaken prior to any further analyses. Descriptive statistics were examined to explore for potential data entry errors and outliers. No outliers or missing data points were found. The data did not deviate significantly from a Gaussian distribution (e.g., skew, kurtosis, non-linearity, range restriction), and compromising levels ($r > .8$) of multicollinearity were not identified. As such, the proposed analyses were utilized for this dat.

Interrater Reliability

An intraclass correlation was conducted based on a mean-rating ($k = 2$), consistency, two-way random-effects model. The intraclass correlation between the two raters in this study was found to be .82, $p < .001$, which is considered good interrater reliability (Koo & Li, 2016).

Descriptive Statistics

Univariate comparisons between age, total years of education, the SHIPLEY-1, and the Familiar Idiom Subscale, Unfamiliar Idiom Subscale, Proverb Subscale, Metaphor Subscale, and

Total Achievement Score are presented in Table 3 below. Age was significantly associated with higher levels of education, and higher scores on the SHIPLEY-1, but not with any of the figurative language scales. Higher levels of education were significantly associated with higher scores on the SHIPLEY-1 and higher scores on the Familiar Idioms Subscale, but not with any other figurative language scales. Scores on the SHIPLEY-1 were positively associated with scores on Familiar Idioms Subscale, Proverbs Subscale, Unfamiliar Idioms Subscale, and the Figurative Language Total Score. Each of the figurative language subscales—Familiar Idioms, Unfamiliar Idioms, Proverbs, Metaphors, and Figurative Language Total—demonstrated significant positive correlations with each other.

Table 3.
Means, Standard Deviations, and Correlations of Sample Characteristics and Main Subscale Score

	M	*SD*	1	2	3	4	5	6	7
1. Age	24.48	8.06							
2. Total Years of Education	14.34	2.01	.57**						
3. Vocabulary	30.17	4.69	.28*	.40**					
4. Familiar Idioms	50.82	4.92	.21	.26*	.42**				
5. Proverbs	36.5	9.72	-.01	.14	.43**	.60**			
6. Metaphors	38.98	8.91	-.21	.08	.21	.31*	.68**		
7. Unfamiliar Idioms	46.68	6.13	.05	.09	.30*	.59**	.51**	.48**	
8. Figurative Language	171.98	24.22	-.03	.16	.41**	.71**	.90**	.83**	.76**

Note: Vocabulary was measured via the SHIPLEY-1 Verbal Reasoning Subtest.
**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Independent samples t-tests examining sex differences, did not indicate any significant differences between male participants' (n = 21) and female participants' (n = 43) performance on the SHIPLEY-1, Familiar Idiom Subscale, Unfamiliar Idiom Subscale, Proverb Subscale, Metaphor Subscale, or Figurative Language Total Score. Additionally, independent samples t-

tests did not indicate any significant differences between monolinguals' (i.e., those who are only fluent in English; $n = 51$) and multilinguals' (i.e., those who are fluent in two or more languages; $n = 13$) performance on the SHIPLEY-1, Familiar Idiom Subscale, Unfamiliar Idiom Subscale, Proverb Subscale, Metaphor Subscale, or Figurative Language Total Score.

The means and standard deviations for each of the four figurative language subscales can be found in the tables below. Table 4 represents the Familiar Idioms subscale, Table 5 represents the Unfamiliar Idioms subscale, Table 6 represents the Proverbs subscale, and Table 7 represents the Metaphors subscale.

Table 4.
Means and Standard Deviations of Familiar Idioms and the Familiar Idiom Subscale Score

	M	*SD*
Item 1	3.79	.64
Item 2	3.89	.29
Item 3	3.80	.58
Item 4	3.61	1.08
Item 5	3.63	.88
Item 6	3.61	.46
Item 7	.95	1.36
Item 8	3.25	1.06
Item 9	3.50	.93
Item 10	3.70	.43
Item 11	3.39	1.16
Item 12	3.27	1.09
Item 13	2.91	1.03
Item 14	3.82	.59
Item 15	3.71	.61
Familiar Idiom Subscale	50.82	4.92

Table 5.
Means and Standard Deviation of Unfamiliar Idioms and the Unfamiliar Idiom Subscale Score

	M	*SD*
Item 16	2.46	1.19
Item 17	3.16	1.07
Item 18	2.36	1.75
Item 19	3.23	1.30
Item 20	3.48	1.01
Item 21	3.41	1.08
Item 22	3.62	.63
Item 23	2.48	1.20
Item 24	3.52	.99
Item 25	2.02	1.71
Item 26	3.54	.89
Item 27	3.30	.82
Item 28	3.18	1.13
Item 29	3.59	.89
Item 30	3.32	1.09
Unfamiliar Idiom Subscale	46.68	6.13

Table 6.
Means and Standard Deviation of Proverbs and the Proverb Subscale Score

	M	*SD*
Item 31	3.09	1.21
Item 32	2.29	1.44
Item 33	2.32	1.61
Item 34	2.30	1.67
Item 35	.64	1.09
Item 36	1.66	1.58
Item 37	2.91	1.09
Item 38	3.36	.90
Item 39	1.66	1.46
Item 40	2.86	1.27
Item 41	2.30	1.56
Item 42	2.86	1.53
Item 43	3.05	1.35
Item 44	1.91	1.71
Item 45	3.29	1.09
Proverb Subscale	36.50	9.72

Table 7.
Means and Standard Deviation of Metaphors and the Metaphor Subscale Score

	M	*SD*
Item 46	3.04	1.17
Item 47	3.04	1.18
Item 48	2.29	1.64
Item 49	2.16	1.41
Item 50	3.27	.81
Item 51	2.09	1.21
Item 52	2.57	1.33
Item 53	2.98	1.05
Item 54	1.59	1.19
Item 55	2.43	1.34
Item 56	2.59	1.29
Item 57	3.27	1.06
Item 58	2.23	1.48
Item 59	2.95	1.35
Item 60	2.50	1.43
Metaphor Subscale	38.98	8.91

Item Reduction Methods

Item Discrimination Analysis. To aid in the reduction of the number of items on the Figurative Language Interpretation Task, an item discrimination analysis was conducted by correlating each item's average score with the average total score on the item's respective subscale as well as the Figurative Language Total Score. The discrimination analysis for each subscale is presented below – Familiar Idioms are presented in Table 8, Unfamiliar Idioms are presented in Table 9, Proverbs are presented in Table 10, and Metaphors are presented in Table 11.

Table 8.
Correlations between Familiar Idioms, the Familiar Idioms Subscale, and the Figurative Language Total Score

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Idiom Subscale
Item 1																
Item 2	-.11															
Item 3	-.02	.08														
Item 4	-.11	.34**	.40**													
Item 5	.04	.11	.06	-.14												
Item 6	.23	-.16	.22	.14	-.13											
Item 7	.10	.11	-.09	.10	.18	-.01										
Item 8	.02	-.07	.41**	.01	.02	.21	-.03									
Item 9	.08	.36**	.13	.11	.07	.00	.15	.23								
Item 10	.14	.15	-.02	.18	.15	.03	.19	-.06	.14							
Item 11	-.07	.06	-.11	.05	-.06	-.01	.08	.03	.19	.02						
Item 12	.07	.08	.25*	.05	.21	.06	.06	.39**	.06	-.05	-.01					
Item 13	.12	-.08	.11	.13	.20	.07	.11	.22	.03	.20	-.01	.01				
Item 14	-.09	.09	.75**	.50**	.13	.18	-.17	.44**	.06	.00	-.12	.36**	.26*			
Item 15	.06	.12	.08	.09	.06	.27*	-.06	.32**	.06	-.05	-.02	-.02	.27*	.49**		
Familiar Idioms Subscale	.21	.28*	.48**	.45**	.33**	.27*	.41**	.55**	.45**	.28*	.26*	.47**	.47**	.58**	.37**	
Figurative Language Total Score	.27*	.06	.27*	.18	.11	.18	.36**	.50**	.52**	.29*	.14	.37**	.29*	.30*	.14	.71**

Note: **Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Table 9.
Correlations between Unfamiliar Idioms, the Unfamiliar Idioms Subscale, and the Figurative Language Total Score

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Unfamiliar Idiom Subscale
Item 16																
Item 17	.20															
Item 18	.04	.20														
Item 19	.00	.06	.32**													
Item 20	.06	.01	.28*	.00												
Item 21	.24	.14	.13	.00	-.16											
Item 22	-.09	-.04	-.05	-.02	.05	-.12										
Item 23	.34**	-.10	.22	.32*	.09	.05	-.12									
Item 24	.11	.16	.16	.00	-.13	.19	-.18	.19								
Item 25	.00	-.11	-.36**	-.20	.00	-.26*	.09	-.18	-.09							
Item 26	.03	.02	-.06	-.18	.24	-.09	-.19	.14	.24	.12						
Item 27	-.08	.00	.20	-.21	.02	.25*	-.05	.08	.34**	.02	.24					
Item 28	.20	.02	.20	.15	.10	.23	-.12	.10	.35**	.09	.03	.24				
Item 29	-.06	-.04	.05	-.09	.13	.01	.04	.32*	.02	.13	.19	.40**	.11			
Item 30	.25*	.22	.00	.07	.09	.36**	-.17	.19	.17	-.02	.35**	.30*	.33**	.01		
Unfamiliar Idioms Subscale	.43**	.32**	.49**	.29*	.32**	.34**	-.06	.49**	.44**	.08	.33**	.44**	.57**	.35**	.55**	
Figurative Language Total Score	.31*	.21	.47**	.07	.44**	.33**	-.05	.27*	.18	.03	.24	.44**	.49**	.29*	.42**	.76**

Note: **Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Table 10.
Correlations between Proverbs, the Proverbs Subscale, and the Figurative Language Total Score

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	Proverb Subscale
Item 31																
Item 32	.08															
Item 33	.12	.26*														
Item 34	.29*	.24	.19													
Item 35	.26*	.15	.15	.33**												
Item 36	.30*	.22	.09	.27*	.17											
Item 37	.33**	.13	.15	.14	.00	.28*										
Item 38	.06	.10	.08	.12	.15	.17	.18									
Item 39	.21	.13	.07	.05	.32**	.21	.26*	.19								
Item 40	.11	.24	-.10	.21	.00	.22	.12	-.11	.07							
Item 41	.26*	.40**	.10	.15	.10	.33**	.06	.20	.09	.31*						
Item 42	.33**	.11	.04	.18	-.11	.17	.18	.09	-.03	.27*	.15					
Item 43	.32*	.35**	.31*	.21	.21	.25*	.09	-.04	.06	.06	.45**	.19				
Item 44	.04	.31*	.08	.18	.17	.21	.15	.03	.09	.20	.17	.10	.37**			
Item 45	.32*	.21	-.06	.31*	.30*	.06	.22	.02	.02	.18	.01	.21	-.08	.11		
Proverbs Subscale	.55**	.58**	.38**	.57**	.43**	.58**	.44**	.28*	.38**	.40**	.56**	.42**	.56**	.49**	.37**	
Figurative Language Total Score	.46**	.52**	.44**	.54**	.39**	.49**	.35**	.19	.26*	.34**	.54**	.43**	.57**	.39**	.31*	.90**

Note: **Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Table 11.
Correlations between Metaphors, the Metaphor Subscale, and the Figurative Language Total Score

	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	Metaphor Subscale
Item 46																
Item 47	.36**															
Item 48	.14	.20														
Item 49	.20	.16	.32*													
Item 50	.13	.42**	.18	.06												
Item 51	.23	.39**	.29*	.31*	.34**											
Item 52	.13	.18	.04	.23	.15	.23										
Item 53	-.01	.36**	.10	.22	.19	.30*	.20									
Item 54	.19	.37**	.46**	.33**	.12	.45**	.15	.19								
Item 55	.05	.28*	.31*	.14	.04	.38**	.38**	.38**	.27*							
Item 56	.20	-.09	.00	.15	-.06	.21	.13	.02	.16	-.20						
Item 57	.13	.02	.11	-.16	-.04	.03	-.03	-.14	.20	.03	-.07					
Item 58	.24	.22	-.02	.08	-.11	-.05	-.03	-.09	.09	.04	.07	.22				
Item 59	.13	.32*	.05	.37**	.14	.21	.28*	.07	.33**	.03	.24	.18	.37**			
Item 60	.07	.39**	.29*	.00	.21	.41**	.21	.17	.18	.41**	.07	.04	-.08	.17		
Metaphor Subscale	.44**	.63**	.53**	.51**	.34**	.66**	.47**	.40**	.64**	.52**	.27*	.20	.30*	.56**	.51**	
Figurative Language Total Score	.28*	.61**	.43**	.44**	.33**	.59**	.36**	.38**	.61**	.56**	.12	.14	.18	.36**	.42**	.83**

Note: **Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed)

The item-to-subscale correlations were assessed to determine which items were eligible to be maintained in the subscale. Items with a Pearson's *r* correlation greater than or equal to 0.2 were considered acceptable items to maintain in the subscale (Schweizer & DiStefano, 2016, pg. 98). Among the familiar idioms, all items (items 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15) demonstrated significant associations with the Familiar Idioms Subscale score, except item 1. Among the unfamiliar idioms, all items (items 16, 17, 18, 19, 20, 21, 23, 24, 26, 27, 28, 29, and 30) demonstrated significant associations with the Unfamiliar Idioms Subscale score, except items 22 and 25. Each of the proverbs (items 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, and 45) demonstrated significant associations with the Proverbs Subscale score. Among the metaphors, all items (items 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 58, 59, and 60) demonstrated significant associations with the Metaphors Subscale score, except item 57.

All significant correlations surpassed the minimum threshold of a Pearson's *r* of 0.2 or higher, which allowed these items to maintain eligibility in the final measure. Additionally, it was decided that the Familiar Idioms subscale demonstrated strong enough psychometric properties to be included in the final measure, and the Unfamiliar Idioms were not included in further analyses or the final measure.

Item Difficulty Index. In addition to the item discrimination analysis, an item difficulty index was created by computing the average score across all participants for each item. Scores near "0" represented more difficult items, while scores nearer to "4" represented easier items. The item difficulty index is presented in Table 12.

Table 12.
Item Difficulty Index for the Figurative Language Interpretation Task Pilot Items

Item Number	Item Difficulty	Item Number	Item Difficulty
35	.64	53	2.98
7	.95	46	3.04
54	1.59	47	3.04
36	1.66	43	3.05
39	1.66	31	3.09
44	1.91	8	3.25
51	2.09	12	3.27
49	2.16	50	3.27
58	2.23	57†	3.27
32	2.29	45	3.29
48	2.29	38	3.36
34	2.3	11	3.39
41	2.3	9	3.5
33	2.32	4	3.61
55	2.43	6	3.61
60	2.5	5	3.63
52	2.57	10	3.7
56	2.59	15	3.71
40	2.86	1†	3.79
42	2.86	3	3.8
13	2.91	14	3.82
37	2.91	2	3.89
38	3.36		

Note: † Indicates that the item was not eligible for inclusion in the final measure due to non-significant correlation on the item discrimination analysis.

Internal Consistency Analysis. Cronbach's alpha was also calculated for each subscale to remove any potential items that dropped the alpha level below the adequate threshold. Items were removed based on the "alpha-if-deleted" statistic in a series of iterations to optimize alpha levels of the subscales. For the purposes of item reduction, Cronbach's alpha of 0.70 or higher was considered adequate internal consistency for an exploratory measure (Hair et al., 2014; Nunnally, 1978). The Familiar Idiom Subscale initially consisted of 15 items (α = .57), however, after removing items that lowered the alpha level of this subscale and considering each item's

difficulty and discrimination analysis, the final Familiar Idioms Subscale consisted of 6 items that represented the highest possible alpha level for this subscale (α = .67). The Proverbs Subscale initially consisted of 15 items (α = .74), however, after reducing items based on the item discrimination analysis and item difficulty index, the final Proverbs Subscale consisted of 9 items, while maintaining a near-identical alpha level (α = .73). Similarly, the Metaphors Subscale initially consisted of 15 items (α = .74), however, after reducing items based on the item discrimination analysis and item difficulty index, the final Metaphors Subscale consisted of 7 items, while maintaining a similar alpha level (α = .72). Together, the original Figurative Language Interpretation Task consisted of 60 items, and after applying the aforementioned item reduction techniques the final Figurative Language Interpretation Task consisted of 22 items, with good internal consistency (α = .83).

Psychometrics of the Final Figurative Language Interpretation Task

In summary, after utilizing an item discrimination analysis, calculating an item difficult index, and obtaining acceptable alpha levels on all scales, the 60-item measure was reduced to a final 22-item Figurative Language Interpretation Task. The final measure consists of 6 idioms, 9 proverbs, and 7 metaphors. Each of the items included correlated significantly with its respective subscale scores (r ranged from .27 to .66) and represented a range of difficulty according to the item difficulty index (ranging from 1.59 to 3.82). The three subscales—the Familiar Idioms Subscale (α = .67), the Proverbs Subscale (α = .73), and the Metaphors Subscale (α = .72)—approached or surpassed the minimum internal consistency threshold of .70 necessary for exploratory measures (Hair et al., 2014; Nunnally, 1978). Additionally, the Figurative Language Interpretation Task demonstrated good internal consistency (α = .83). The final psychometric

properties of the 22-item Figurative Language Interpretation Task are presented in Table 13 below.

Table 13.
Psychometrics of Revised Figurative Language Measure

Item	Subscale	Item Difficulty	Correlation with Respective Subscale	Correlation with Total Score
3	Idiom	3.80	.48**	.27*
6	Idiom	3.61	.27*	.18
8	Idiom	3.25	.55**	.50**
13	Idiom	2.91	.47**	.50**
14	Idiom	3.82	.58**	.30*
15	Idiom	3.71	.37**	.14
31	Proverb	3.10	.55**	.46**
32	Proverb	2.29	.58**	.52**.
34	Proverb	2.30	.57**	.54**
36	Proverb	1.66	.58**	.49**
40	Proverb	2.86	.40**	.34**
41	Proverb	2.30	.56**	.54**
43	Proverb	3.05	.56**	.57**
44	Proverb	1.91	.49**	.39**
45	Proverb	3.29	.37**	.31*
47	Metaphor	3.04	.63**	.61**
48	Metaphor	2.29	.53**	.43**
50	Metaphor	3.27	.34**	.33**
51	Metaphor	2.09	.66**	.59**
53	Metaphor	3.00	.40**	.38**
54	Metaphor	1.59	.64**	.61**
55	Metaphor	2.43	.52**	.56**

Note: **Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Study Two

Method

Participants

A power analysis conducted with GPower was based on the statistical analysis which required the largest number of participants (i.e., analyses required to test hypothesis 2). The power analysis determined that study two would require 100 participants to achieve 80% power. Accordingly, a total of 100 community members and students from a midsized university in the Midwestern U.S. participated in this study. Community members did not receive compensation for participation; however, students were compensated with either required research credits for a course or extra credit. Participants were recruited through the SONA Experiment Management System, an online participant management software system, and through advertisements distributed across the university's campus. Individuals were eligible to participate if they were fluent American-English speakers and were between the ages of 18 and 89 years old. However, participants with a history of severe psychiatric disorder (i.e., Schizophrenia, Bipolar Disorder), neurodevelopmental disorder (e.g., developmental disorder, learning disorder), or brain injury known to impact their cognition were excluded. The total sample (N = 100) was composed of 32 male participants (32%) and 68 female participants (68%). The participants' average age was 22.96 (SD = 8.23) years. There were 69 White/non-Hispanic (69%), 8 African-American/Black (8%), 11 Hispanic/Latino (11%), 8 Asian/Pacific Islander (8%), 1 American Indian/Native Alaskan/Aleutian/Eskimo (1%), and 3 bi-racial (3%) participants. All participants were fluent in American English, however, 90 participants (90%) indicated that English was their primary or native language, while 10 participants (10%) indicated that English was not their primary or native language. Seventy-six participants (76%) reported being fluent in only one language, 22

participants (22%) reported being fluent in two languages, and 2 participants (2%) reported being fluent in three or more languages. On the Demographic Index of Cultural Exposure (DICE; Cruz et al., 2012), six is the maximum score, which indicates a high degree of acculturation to American culture, 68 participants (68%) received a score of 6, 11 participants (11%) received a score of 5, 11 participants (11%) received a score of 4, 4 participants (4%) received a score of 3, 2 participants (2%) received a score of 2, 2 participants (2%) received a score of 1, and 2 participants (2%) received a score of 0.

Procedures

Ethical approval was obtained from the Wichita State University Institutional Review Board. The assessments in this study were administered by a graduate student and two undergraduate research assistants who were trained on the standardized administration of each instrument. Each research assistant administered three batteries of the testing procedures under the observation of the graduate student before administering the tests independently. Periodically, the researchers observed each other's sessions to ensure adherence to standardization protocols.

Each session was conducted in-person, and individually with each participant. First, participants were consented into the study, and then the assessment instruments were administered in the following order: Digit Span, Similarities, the Stroop Task, the Wisconsin Card Sorting Test, the Figurative Language Interpretation Task, the SHIPLEY-2, the TIPI, and the demographics form. For each task, the examiner read aloud the instructions and answered any questions according to the standardization procedures for each instrument. Relevant data was recorded on a data collection form by the examiner. Each session of the study took approximately 60 minutes to complete.

Materials

Demographics Form. The demographics form was a self-report questionnaire made by the author. This form collected information on age, sex, race/ethnicity, educational level, and total years of education in accordance with the American Psychology Association's guidelines for assessing multicultural demographic information (American Psychological Association [APA], 2017; APA, 2019). Participants were also asked to report on their native language, number of languages the participant is fluent in, their country of origin, and their history of speech and language, learning, or other developmental disorders.

In addition to this data, the demographics form utilized questions from the Demographic Index of Cultural Exposure (DICE; Cruz et al., 2012) to assess the degree of cultural exposure that each participant has to the majority culture (i.e., United States culture). In calculating the DICE, a value of "1" was assigned to each of the following indicators to reflect U.S. cultural exposure, while a value of "0" was assigned if an individual did not have exposure to that indicator. The six indicators include (1) country where the participant attained highest level of formal education, (2) country where the participant lived the longest, (3) participant's native language, (4) participant's country of birth, (5) participant's mother's country of birth, and (6) participant's father's country of birth. The six items were then summed to obtain a DICE score for each participant. When compared to other measures of acculturation, such as the Cultural Lifestyles Inventory (CLSI, Mendoza, 1989), the Bidimensional Acculturation Scale (BAS; Marin & Gamba, 1996), and fatalism and folk illness belief scales (Cuellar et al., 1995), it was found that the DICE successfully differentiates individuals along a spectrum of cultural exposure, and that an individual's index score is associated with their cultural beliefs, behaviors, and preferences (Cruz et al., 2012).

Ten-Item Personality Inventory. The Ten-Item Personality Inventory (TIPI; Gosling et al., 2003) is a paper-and-pencil self-report questionnaire that is composed of 10 items designed to assess the Big Five personality dimensions of Agreeableness, Conscientiousness, Emotional Stability, Extraversion, and Openness to Experience. Each Big Five personality domain is assessed through two items, one coded in the positive direction and one in the negative direction. Each item is composed of the statement "I see myself as:" followed by two descriptive words or phrases. The responses to the statement range from "disagree strongly" to "agree strongly" rated along a 7-point Likert scale. In previous research, the TIPI has demonstrated good test-retest reliability over a six-week period (mean r = .72), with subscale test-retest reliability ranging from .62 (Openness to Experience) to .77 (Extraversion; Gosling et al., 2003). As this measure includes only two items per sub-scale, it is not designed to have a high alpha level. The TIPI has demonstrated convergent validity with other measures of Big Five personality domains, including the Five Factor Form and the Big Five Inventory (John et al., 1991; Rojas & Widiger, 2014; Gosling et al., 2003).

SHIPLEY Institute of Living Scales – Second Edition. The SHIPLEY Institute of Living Scales—Second Edition (Shipley-2; Shipley et al., 2009) is a brief measure of crystallized abilities (abilities gained through education and experience) and fluid abilities (the ability to problem solve or learn and acquire new information; Kaya et al., 2012). The Vocabulary scale is used to measure crystallized ability, while either the Block Patterns or Abstraction scale is used to measure fluid abilities. The Vocabulary and Abstraction scales form Composite A, while the Vocabulary and Block Pattern scales form Composite B. Composite A (Vocabulary and Abstraction) was used in the present study. The Vocabulary scale includes 40 items. Each item prompts the participant to read a target word and then choose from four options which word is

closest in meaning to the target word. There is one correct answer, and the other three options are foils. The Abstraction scale includes 25 items. Each item includes a sequence or pattern, and the participant is required to fill in what they believe accurately completes or continues the sequence. The SHIPLEY-2 takes approximately 20-25 minutes to administer. Evidence for convergent validity has been demonstrated through positive correlations with other measures of intelligence, such as the and the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III; Wechsler, 1997; Shipley et al., 2009), and Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV; Wechsler, 1997; Lodge, 2013). The test-retest correlation coefficients ranged from .74 to .94 for the subscales and from .76 to .94 for the resulting composite scores (Shipley et al., 2009). Internal consistency reliability estimates for ranged from .88 to .97 with a median of .91 (Shipley et al., 2009).

The Stroop Color and Word Test (SCWT). The Stroop Color and Word Test (SCWT; Stroop, 1935) is a neuropsychological test used to assesses the ability to maintain a goal in mind, selectively attend to different stimuli, and inhibit automatic responses in favor of a more effortful, or controlled responses (Strauss et al., 2006). The Victoria Stroop Test (VST; Regard, 1981 cited in Strauss et al., 2006) was utilized in the present study, which takes approximately 5 minutes to administer. The VST includes three white cards with different stimuli printed on them—one page with colored dots (green, red, blue, and yellow), one page with names of colors (green, red, blue, and yellow) printed in black ink, and one page with the names of the colors printed in an incongruent colored ink (e.g., the word “red” printed in green ink). Each card contains 24 of these stimulus items. Of primary interest is the participant’s behavior when presented with colored words printed in incongruent colored inks, as the participants are required to perform a less automatic task (i.e., naming ink color) while inhibiting the interference from a

more automatic task (i.e., reading the word; MacLeod & Dunbar, 1988; Ivnik et al., 1996). This difficulty in inhibiting the more automatic process is called the Stroop effect (Stroop, 1935), with slower completion times and higher number of errors representing more difficulty with inhibition (Strauss et al., 2006). The Completion Time score and the Number of Errors score will be used for the current study. The Stroop interference score has been related to another measure of response inhibition—the difference score between Trails A and B (r = .55; May & Hasher, 1998). Additionally, Stroop interference loads on the same factors as several WAIS-III measures of processing speed, conceptual abilities, and attention (i.e., Digit Symbol, Block Design, Similarities, Digit Span; Graf et al.,1995; Bondi et al., 2002). Bullock et al. (unpublished data) retested participants on the VST with a one-month interval between test sessions. Reliability coefficients were high—.90, .83, and .91 for each of the three parts of the test (cited in Strauss et al., 2006). However, participants demonstrated significant practice effects, improving by several seconds on the interference trial. Correlations among test trials tend to be moderate in normal individuals for the VST (Pineda & Merchan, 2003).

Wisconsin Card Sorting Test. The Wisconsin Card Sorting Test (WCST; Berg, 1948; Grant & Berg, 1948) is a measure of planning, goal-oriented behavior, inhibition of impulsive responses, and the ability to form abstract concepts (Heaton et al., 1981; Heaton et al., 1993). The WCST consists of cards with different shapes (crosses, circles, triangles, or stars) printed in various colors (red, blue, yellow, or green), and arranged in sets of numbers (1, 2, 3 or 4). To begin, four stimulus cards are placed in front of the subject, the first with a red triangle, the second with two green stars, the third with three yellow crosses, and the fourth with four blue circles on them. The participant is then given a stack of cards and is instructed to sort each response card onto one of the stimulus cards in whichever manner she or he thinks is correct. No

other instructions are given throughout the test. After each sort, the participant is told whether the sort was right or wrong. After 10 consecutive correct responses, the sorting rule changes without any warning to the participant. The sequence of correct sorting follows this order: color, shape, number, color, shape, and number. The test ends when the participant has successfully completed six sets of sorting rules, or until all the response cards are used. Typical administration takes 15-20 minutes. The Total Number of Correct Responses and the Number of Perseverative Responses are the two scores that will be utilized for the present study. Total Number of Correct Responses indicates how well the participant understood and executed the task properly, while the number of Perseverative Responses indicates an inability to move from the old category for the new one, or the inability think flexibility about a new possibility (Strauss et al., 2006). The WCST has been found to modestly correlate with other measures of general reasoning, such as Full-Scale IQ (Ardila et al., 2000; Greve et al., 1999; Sherman et al., 1995). Additionally, performance on the WCST appears to measure conceptual processing similar, but not identical, to that of other tests of rule derivation such as the Weigl Test (Laiacona et al., 2000) and the Category Test (e.g., Golden et al., 1997; Pendleton & Heaton, 1982). Scores on WCST appear to measure attention and working memory, as WCST has been found to load onto similar factors as scores from Trails B, Continuous Performance Test, and Digit Symbol (Pukrop et al., 2003, Greve et al., 1998, Greve et al., 1999). In various studies the WCST has demonstrated poor test-retest reliability, as the test appears to be sensitive to practice effects (Basso et al., 1999; Paolo et al., 1996; Ferland et al., 1998).

Digit Span (from the Wechsler Adult Intelligence Scale – Third Edition). The Digit Span (DS) subtest of the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III; Wechsler 1997) includes two tasks—digit span forward and digit span backward. Digit span forward is a

measure of attention span and short-term memory, while digit span backward is a measure of working memory (Wechsler, 1997). DS forward involves asking participants to repeat a series of digits verbatim. The task starts with two digits and increases by one digit every two trials. DS backward involves asking participants to repeat digits in the reverse order of presentation. DS forward and backward takes approximately 10 minutes to administer. Practice effects on these tasks are negligible and there is high test-retest reliability when measured across various weeks (r = .90; Psychological Corporation, 1997; McCaffery et al., 2000). The DS subtest has demonstrated high correlations with the Wechsler Intelligence Scale for Children—Third Edition Digit Span subtest (r = .73), and with the Stanford-Binet Intelligence Scale—Fourth Edition Short-Term Memory scale (r = .52; Psychological Corporation, 1997).

Measure of Effort: Reliable Digit Span. Reliable Digit Span (RDS) is one of the oldest and most heavily researched symptom validity tests and is commonly used in clinical practice to detect suspect effort (Schroeder et al., 2012). The RDS is calculated by "summing the longest string of digits repeated without error over two trials under both forward and backward conditions" (Greiffenstein et al., 1994, pp. 219-220). A meta-analysis of over 20 studies on RDS has indicated that a cutoff score of ≥ 6 can be used effectively in a range of samples with specificity rates between 96-97% (Schroeder et al., 2012), and as such, scores ≥ 6 are considered evidence that the participant gave sufficient effort during testing.

Similarities (from the Wechsler Adult Intelligence Scale – Third Edition). The Similarities subtest of the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III; Wechsler 1997) is a measure of verbal abstract reasoning. Similarities contains 19 items, which are pairs of words (e.g., "piano" and "drum"). The examiner reads the word pair to the examinee, and the examinee is tasked with describing how the two words (either objects or intangible

concepts) are alike (Wechsler, 1997). Similarities takes approximately 10 minutes to administer. Similarities requires clinical judgment in scoring, however, inter-rater reliability for Similarities is good, with a coefficient of .93 (Psychological Corporation, 1997). Test-retest reliability measured across several weeks for Similarities is also good, with a coefficient of .88 (Psychological Corporation, 1997). The Similarities subtest is a component of the Verbal Comprehension Index, which has demonstrated high correlations with the Stanford-Binet Intelligence Scale—Fourth Edition verbal reasoning scale (r = .87; Psychological Corporation, 1997).

Figurative Language Interpretation Task. The Figurative Language Interpretation Task consisted of 22 sayings that were presented in a free inquiry format—6 idioms, 9 proverbs, and 7 metaphors. The same administration and scoring protocol described for study one was utilized for study two. Additionally, the same training protocol used in study one was utilized in study two to train both the primary and secondary raters of the responses on the Figurative Language Interpretation Task.

Data Analysis

Preparation of the Data. Following data cleaning, descriptive statistics were run for demographic variables, personality traits, scores on the Figurative Language Interpretation Task, and scores on each cognitive measure. Additionally, univariate comparisons and ANOVAs utilizing demographic variables, personality traits, scores on the Figurative Language Interpretation Task, and scores on the cognitive measures were run to examine potential differences and associations that may influence subsequent analyses and interpretation of the data.

Furthermore, to ensure interrater reliability of the figurative language interpretation scores a random sampling of 20 participant protocols was scored by a second rater and an intraclass correlation was conducted between the two raters' scores. According to Koo and Li, (2016), "values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability."

Determining the Psychometric Properties of the Figurative Language Interpretation Task. The same psychometric analyses as were conducted in study one were also conducted on the figurative language scores in study two to further determine the characteristics of the Figurative Language Interpretation Task. Correlations were conducted between each item and its respective scale, and between each item and Figurative Language Total Score. An item difficulty index was also calculated by computing the average score across all participants for each item. Scores near "0" represent more difficult items, while scores near "4" represent easier items. Additionally, a Spearman Rho rank order correlation was conducted to assess the relationship between scores on each item in study one and study two. Furthermore, to address aims (1) and (2) of study two, an exploratory factor analysis was conducted with a minimum average partial test to determine the number of factors (Caron, 2019). For any identified scales and for the figurative language subscale scores, Cronbach's alpha was conducted to assess internal consistency of scales on the Figurative Language Interpretation Task. For exploratory measures, a minimum internal consistency threshold of .70 is recommended (Hair et al., 2014; Nunnally, 1978).

Hypothesis Testing. Hypothesis (1) stated that the three types of figurative language would demonstrate a positive association with other measures of abstract reasoning. To address

this hypothesis, a canonical correlation was conducted with the three figurative language subscale scores (the Idioms Subscale, the Proverbs Subscale, and the Metaphors Subscale) entered as the independent variables, while the measures of verbal abstract reasoning (as measured by Similarities) and abstract reasoning (as measured by SHIPLEY-2 Abstraction) were input as the dependent variables.

Hypothesis (2) predicted that performances on the idiom, proverb, and metaphor subscales would significantly contribute to predicting abstract reasoning performance. To address this hypothesis two forced entry linear regressions were conducted. One regression utilized verbal abstract reasoning (as measured by Similarities) as the outcome variable. The Idiom Subscale, the Proverb Subscale, and the Metaphor Subscale were entered into the regression, along with significant associations between demographic variables and Similarities in the univariate comparisons. The second regression utilized abstract reasoning (as measured by SHIPLEY-2 Abstraction) as the outcome variable. The Idiom Subscale, the Proverb Subscale, and the Metaphor Subscale were entered into the regression, along with significant associations between demographic variables and SHIPLEY-2 Abstraction in the univariate comparisons.

Hypothesis (3) was that the idioms, proverbs, metaphors, and the figurative language total score would each demonstrate a positive association with a) intelligence scores, b) inhibition scores, c) working memory scores, and d) cognitive flexibility scores. To address hypothesis (3) a series of Pearson's *r* correlations were conducted between the Idioms Subscale, the Proverbs Subscale, the Metaphors Subscale, Figurative Language Total Score, and scores on measures of cognition. Specifically, to address hypothesis (3a), involving higher intelligence, correlations were conducted between figurative language scores and the SHIPLEY-2 Composite Scale. For hypothesis (3b), involving inhibition ability, correlations were conducted between

figurative language scores and the Stroop Color Word Test scores. For hypothesis (3c), involving working memory, correlations were conducted between figurative language scores and the Digit Span Backward score. For hypothesis (3d), involving cognitive flexibility, correlations were conducted between figurative language scores and the Wisconsin Card Sorting Test scores.

Hypothesis (4) was that male participants and female participants would vary in their figurative language interpretation abilities. To examine any potential sex differences, male participants' and female participants' scores on the figurative language subscales (the Idioms Subscale, the Proverbs Subscale, the Metaphors Subscale, and the Figurative Language Total Score) were examined using a series of between-groups ANOVAs.

Model 4, below, summarizes the relationships between figurative language interpretation and cognitive abilities that will be explored in study two, while Table 14 summarizes the instruments employed in study two, along with the scores used from each instrument, and the psychological construct which each score represents.

Model 4.
Relationships between Figurative Language Interpretation and Cognitive Abilities that will be examined in Study Two

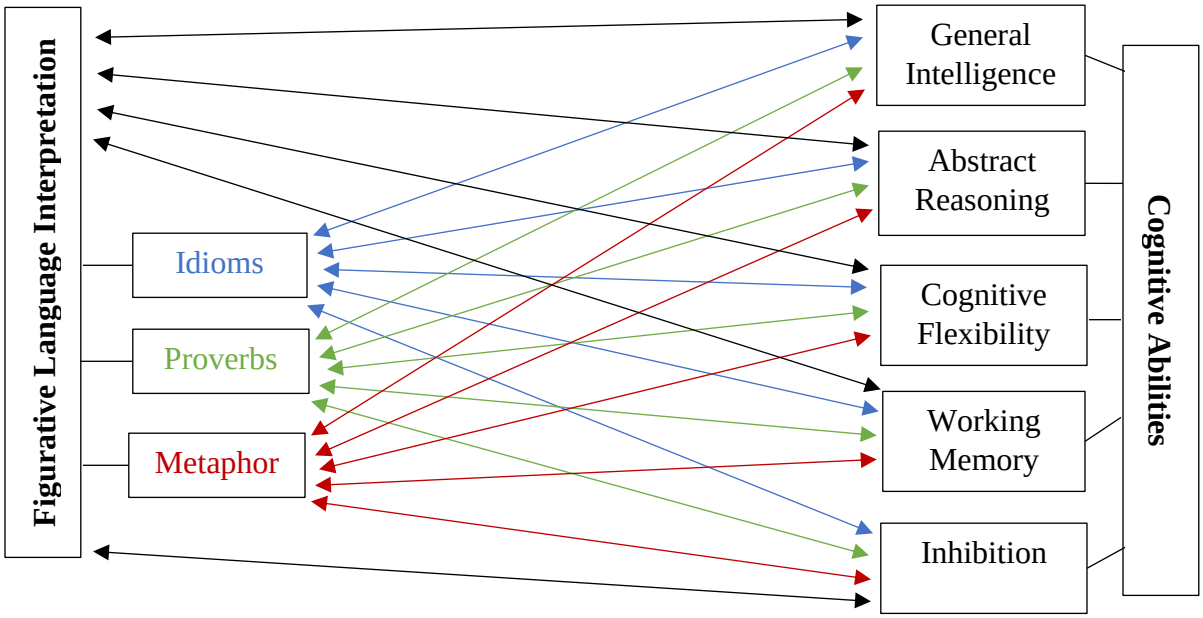


Table 14.
Summary of Measures, Scores, and Associated Constructs for Study Two

Test	Scores Used	Construct Measured
Figurative Language Interpretation Task	Idioms Subscale Score	Idiom interpretation
	Proverbs Subscale Score	Proverb interpretation
	Metaphors Subscale Score	Metaphor interpretation
	Total Score	Figurative language interpretation
SHIPLEY-2	Vocabulary Score	Crystallized knowledge
	Abstraction Score	Fluid (abstract) reasoning
	Composite A Score	General intelligence
Similarities	Total Score	Verbal abstract reasoning
Stroop Color and Word Test	Condition 3 Total Time	Inhibition
	Condition 3 Total Errors	Inhibition
Wisconsin Card Sorting Test	Total Number of Correct Responses	Cognitive flexibility
	Perseverative Responses	Cognitive flexibility
Digit Span	DS Backward Score	Working memory

Results

Data Cleaning

After data collection was completed, data cleaning procedures were undertaken prior to any hypothesis testing. Descriptive statistics were examined to explore for potential data entry errors and outliers. No outliers were found, and only two missing data points were discovered within the database. Both these data points were from the same participant who was not administered one trial of an assessment, and their values were replaced with the sample's mean score for those items. Additionally, reliable digit span (RDS) was computed as a measure of valid performance on cognitive testing. All participants met or exceeded the minimum threshold for valid performance according to RDS (Greiffenstein et al., 1994, pp. 219-220, Schroeder et al., 2012), indicating that sufficient effort was given on cognitive testing by all participants. The data did not deviate significantly from a Gaussian distribution (e.g., skew, kurtosis, non-linearity, range restriction), and compromising levels ($r > .8$) of multicollinearity were not identified. As such, the proposed parametric tests were utilized for the analysis of this data.

Interrater Reliability

To ensure accurate scoring of the Figurative Language Interpretation Task, a random sampling of 20 participant protocols was scored by a second rater following the procedures outlined in study one. An intraclass correlation was conducted based on a mean-rating ($k = 2$), consistency, two-way random-effects model. The intraclass correlation between the two raters in this study was found to be .92, $p < .001$, which is considered excellent reliability (Koo & Li, 2016).

Descriptive Statistics

Descriptive statistics were run for demographic variables, personality traits, scores on the Figurative Language Interpretation Task, and scores on each cognitive measure. The means and standard deviations of these variables are displayed below, with the demographic variables and Big Five personality traits displayed in Table 15, the Figurative Language Interpretation Task displayed in Table 16, and each cognitive measure displayed in Table 17.

Table 15.
Means and Standard Deviation of Demographic and Personality Variables

	M	*SD*
Age	22.96	8.23
Total Years of Education	13.75	1.91
Acculturation	5.25	1.40
Extraversion	7.48	3.08
Agreeable	10.06	2.29
Conscientious	10.90	2.50
Emotionally Stable	8.42	2.65
Openness	10.36	2.21

Note: Acculturation was measured via the Demographic Index of Cultural Exposure.

Table 16.
Means and Standard Deviation of Figurative Language Scores

	M	*SD*
Idioms Subscale	21.17	5.89
Proverbs Subscale	19.55	7.12
Metaphors Subscale	14.66	5.81
Figurative Language Total Score	54.70	12.40

Table 17.
Means and Standard Deviations of Cognitive Testing Scores

	M	*SD*
Working Memory	6.82	2.03
Inhibition (Time)	20.13	4.83
Inhibition (Errors)	.62	1.10
Cognitive Flexibility (Correct)	49.36	7.75
Cognitive Flexibility (Perseverative)	7.71	7.90
Vocabulary	29.08	3.94
Abstract Reasoning	15.40	3.24
IQ	44.48	6.02
Verbal Abstract Reasoning	24.27	3.83

Note: Working Memory was measured via Digit Span Backwards; Inhibition was measured via Stroop Color-Word Test Condition 3 Time and Condition 3 Errors; Cognitive Flexibility was measured via Wisconsin Card Sorting Test Total Correct and Perseverative Errors; Vocabulary was measured via SHIPLEY-2 Vocabulary; Abstract Reasoning was measured via SHIPLEY-2 Abstraction; IQ was measured via SHIPLEY-2 Composite Score; and Verbal Abstract Reasoning was measured via Similarities.

Univariate Comparisons

Correlations were conducted between demographics variables and the Figurative Language Interpretation Task subscales, between demographic variables and cognitive assessment scores, and between personality traits and the Figurative Language Interpretation Task subscales. Additionally, a series of ANOVAs were conducted to examine any significant differences among categorical demographic variables (i.e., sex, English as a native language, and number of languages spoken) and Figurative Language Interpretation Task scores.

Demographics and Figurative Language Comparisons. The univariate comparisons between age, total years of education, acculturation, and the Idiom Subscale, Proverb Subscale, Metaphor Subscale, and Figurative Language Total Score are presented in Table 18 below.

Table 18.
Means, Standard Deviations, and Correlations of Sample Characteristics and Figurative Language Scores

	1	2	3	4	5	6
1. Age						
2. Total Years of Education	.58**					
3. Acculturation	.22*	.20				
4. Idioms	.32**	.21*	.15			
5. Proverbs	.22*	.29**	.26**	.28**		
6. Metaphors	.06	.09	.21*	.08	.41**	
7. Figurative Language	.21*	.29**	.32**	.33**	.87**	.76**

Note: Acculturation was measured via the Demographic Index of Cultural Exposure.
**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Age demonstrated a significant positive association with years of education, acculturation, the Idioms Subscale, the Proverbs Subscale, and the Figurative Language Total Score, but not with the Metaphors Subscale. Higher levels of education demonstrated a significant positive association with the Idioms Subscale, the Proverbs Subscale, and the Figurative Language Total Score, but not with the Metaphors Subscale. Acculturation demonstrated a significant positive association with the Proverbs Subscale, the Metaphors Subscale, and the Figurative Language Total Score, but not with the Idioms Subscale. Each of the figurative language subscales—Familiar Idioms, Unfamiliar Idioms, Proverbs, Metaphors, and Figurative Language Total—demonstrated significant positive correlations with each other, however, the Metaphors Subscale and Idioms Subscale were not significantly correlated.

A series of one-way between-groups analysis of variance (ANOVAs) were conducted to examine any potential differences between monolinguals' (n = 76) and multilinguals' (n = 24) scores on the figurative language subscale scores. There was a statistically significant difference in scores on the Metaphors Subscale for monolinguals and multilinguals: $F(1, 98) = 4.17$, p =

.04. The monolinguals' score (M = 15.32, SD = 5.54) was significantly higher than the multilinguals' score (M =12.58, SD = 6.26). No statistically significant difference was found between monolinguals and multilinguals for the Idioms Subscale ($F(1, 98) = 0.27$, $p = .60$), or the Proverbs Subscale ($F(1, 98) = 1.94$, $p = .17$).

Additionally, a series of between-groups ANOVAs was conducted to explore for potential differences between native English speakers' (n = 90) and nonnative English speakers' (n = 10) scores on the figurative language subscales. No statistically significant differences were found between native English speakers' and nonnative English speakers' scores on the Idioms Subscale ($F(1, 98) = 1.98$, $p = .16$), the Proverbs Subscale ($F(1, 98) = 2.81$, $p = .10$), or the Metaphors Subscale ($F(1, 98) = 0.80$, $p = .37$).

To examine any potential sex differences, male participants' (n = 32) and female participants' (n = 68) scores on the figurative language subscales were examined using a series of one-way between-groups ANOVAs. No statistically significant differences were found between male participants' and female participants' scores on the Idioms Subscale ($F(1, 98) = 0.00$, $p = .99$), the Proverbs Subscale ($F(1, 98) = 1.80$, $p = .18$), or the Metaphors Subscale ($F(1, 98) = 2.22$, $p = .14$).

Personality Traits and Figurative Language Comparisons. In addition to examining demographic variables, personality traits (based on the Big Five model of personality) were included to control for any potential impact on Figurative Language Interpretation Task scores. Correlations were conducted between all five personality traits, the Idioms Subscale, the Proverbs Subscale, the Metaphors Subscale, and the Figurative Language Total Score. These correlations are presented in Table 19 below.

Table 19.
Means, Standard Deviations, and Correlations of Personality Traits and Main Subscale Score

	1	2	3	4	5	6	7	8
1. Extraversion								
2. Agreeable	.11							
3. Conscientious	.09	.30**						
4. Emotionally Stable	.22*	.15	.11					
5. Openness	.19	.18	-.02	.18				
6. Idioms	.11	.08	.08	.00	.06			
7. Proverbs	.26**	.14	.07	.08	.00	.28**		
8. Metaphors	.12	.15	.04	-.03	-.05	.08	.41**	
9. Figurative Language	.20*	.14	.07	.04	-.05	.33**	.87**	.76**

Note: **Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Extraversion was significantly positively associated with the Proverbs Subscale and the Figurative Language Total Score, but not with the Idioms Subscale or the Metaphors Subscale. No other personality traits—agreeableness, conscientiousness, emotional stability, or openness—were significantly associated with any of the figurative language scores.

Demographics and Cognitive Testing Comparisons. A series of correlations were conducted to examine any potential relationships between demographic variables and scores on cognitive testing. These correlations are presented in Table 20 below.

Table 20.

Correlations between Demographics and Cognitive Testing

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Age											
2. Total Years of Education	.58**										
3. Acculturation	.22*	.20									
4. Working Memory	-.13	-.08	.01								
5. Inhibition (Time)	.27**	.01	-.08	-.22*							
6. Inhibition (Errors)	-.01	-.02	-.10	-.16	.30**						
7. Cognitive Flexibility (Correct)	.08	.16	.18	-.08	-.08	.02					
8. Cognitive Flexibility (Perseverative)	-.10	-.16	-.12	.07	.03	.01	-.96**				
9. Vocabulary	.35**	.43**	.45**	-.03	.06	-.08	.14	-.13			
10. Abstract Reasoning	.26**	.22*	.27**	.18	-.10	-.10	.29**	-.26**	.40**		
11. IQ	.37**	.40**	.44**	.08	-.02	-.11	.24*	-.23*	.87**	.80**	
12. Verbal Abstract Reasoning	.23*	.30**	.32**	.17	-.04	-.14	.13	-.16	.64**	.38**	.63**

Note: Acculturation was measured via the Demographic Index of Cultural Exposure; Working Memory was measured via Digit Span Backwards; Inhibition was measured via Stroop Color-Word Test Condition 3 Time and Condition 3 Errors; Cognitive Flexibility was measured via Wisconsin Card Sorting Test Total Correct and Perseverative Errors; Vocabulary was measured via SHIPLEY-2 Vocabulary; Abstract Reasoning was measured via SHIPLEY-2 Abstraction; IQ was measured via SHIPLEY-2 Composite Score; and Verbal Abstract Reasoning was measured via Similarities.

**Correlation is significant at the 0.01 level (2-tailed).

*Correlation is significant at the 0.05 level (2-tailed).

Age demonstrated a significant positive association with inhibition (time), vocabulary, abstract reasoning, IQ, and verbal abstract reasoning. Age was not associated with working memory, inhibition (errors), cognitive flexibility (total), or cognitive flexibility (errors). Years of education demonstrated a significant positive association with vocabulary, abstract reasoning, IQ, and verbal abstract reasoning. Years of education was not associated with working memory, inhibition (time), inhibition (errors), cognitive flexibility (total), or cognitive flexibility (errors). Acculturation demonstrated a significant positive association vocabulary, abstract reasoning, IQ, and verbal abstract reasoning. Acculturation was not associated with working memory, inhibition (time), inhibition (errors), cognitive flexibility (total), or cognitive flexibility (errors).

Psychometrics of the Figurative Language Interpretation Task

The same psychometric analyses conducted in study one were also conducted on the figurative language scores in study two to further determine the characteristics of the Figurative Language Interpretation Task. Correlations were conducted between each item and its respective scale, and between each item and Figurative Language Total Score. An item difficulty index was also calculated by computing the average score across all participants for each item. Scores near "0" represent more difficult items, while scores near "4" represent easier items. Additionally, a Spearman Rho rank order correlation was conducted to assess the relationship between scores on each item in study one and study two. Scores on each item in study one and study two demonstrated a significant positive association with one another ($\rho = .89$, $p < .001$). The psychometric properties of the Figurative Language Interpretation Task utilized in study two are presented in Table 21 below. The items on this measure represented a range of difficulty, and each item correlated significantly with its respective subscale and with the Figurative Language Total Score.

Table 21.
Psychometrics of Figurative Language Interpretation Task from Study Two

Item # (Item # in Study 1)	Subscale	Item Difficulty (*M*)	Standard Deviation (*SD*)	Correlation with Respective Subscale	Correlation with Total Score
1 (3)	Idiom	3.61	.98	.35**	.41**
2 (6)	Idiom	3.48	.95	.23*	.34**
3 (8)	Idiom	3.13	1.26	.43**	.40**
4 (13)	Idiom	2.82	.76	.35**	.31**
5 (14)	Idiom	3.79	.73	.20*	.28**
6 (15)	Idiom	3.66	.59	.23*	.20*
7 (31)	Proverb	3.24	1.23	.37**	.37**
8 (32)	Proverb	1.75	1.55	.54**	.56**
9 (34)	Proverb	2.62	1.70	.51**	.44**
10 (36)	Proverb	1.28	1.58	.50**	.36**
11 (40)	Proverb	2.36	1.45	.54**	.45**
12 (41)	Proverb	1.96	1.66	.61**	.54**
13 (43)	Proverb	2.17	1.69	.60**	.50**
14 (44)	Proverb	1.28	1.53	.61**	.52**
15 (45)	Proverb	2.89	1.46	.31**	.25*
16 (47)	Metaphor	2.44	1.55	.54**	.40**
17 (48)	Metaphor	1.72	1.86	.48**	.32**
18 (50)	Metaphor	2.94	1.19	.45**	.28**
19 (51)	Metaphor	2.24	1.32	.57**	.47**
20 (53)	Metaphor	1.68	1.78	.57**	.51**
21 (54)	Metaphor	1.48	1.31	.56**	.49**
22 (55)	Metaphor	2.16	1.60	.66**	.44**

Note: **Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Analysis of Scales

To address aims (1) and (2), an exploratory factor analysis was conducted with a minimum average partial test to determine the number of factors (Caron, 2019). Upon conducting this analysis, an eight-factor solution emerged, however, the factor loadings were low, unstable, and variable. Several proverbs appeared to emerge as the most useful factor, however, a clear pattern was not discernible. As such, no clear scales were identifiable.

Cronbach's alpha was conducted for the three figurative language subscales and the Figurative Language Total Score. In study two, the three subscales—the Idioms Subscale (α =

.54), the Proverbs Subscale (α = .65), and the Metaphors Subscale (α = .60)—did not meet the minimum internal consistency threshold of .70 recommended for exploratory measures (Hair et al., 2014; Nunnally, 1978). However, the total Figurative Language Interpretation Task total score demonstrated good internal consistency (α = .76).

Hypothesis Testing

Hypothesis One. Hypothesis (1) was that the figurative language scales would demonstrate a positive association with other measures of abstract reasoning (i.e., Similarities and SHIPLEY-2 Abstraction). Three of the figurative language scores, the Proverbs Subscale (r = .52, p < .001), the Metaphors Subscale (r = .32, p = .001), and the Figurative Language Total Score (r = .51, p < .001) demonstrated a significant positive correlation with a measure of verbal abstract reasoning, while the Idioms Subscale did not. Each of the figurative language scores, the Idioms Subscale (r = .25, p = .01), the Proverbs Subscale (r = .51, p < .001), the Metaphors Subscale (r = .25, p = .01), and the Figurative Language Total Score (r = .49, p < .001) demonstrated a significant positive association with a measure of abstract reasoning.

To further address this hypothesis, a canonical correlation was conducted with the three figurative language subscale scores entered as the independent variables, while the measures of verbal abstract reasoning and abstract reasoning were input as the dependent variables. Tests of dimensionality for the canonical correlation analysis, as shown in Table 22, indicate that one of the two canonical dimensions was statistically significant (p < .001). Dimension 1 demonstrated a canonical correlation of .63 between the sets of variables. Table 23 presents the standardized canonical coefficients for the first dimension across both independent and dependent variables. For the independent variables, the canonical correlation was most strongly influenced by proverbs (standardized canonical coefficient of -.88), followed by metaphors (standardized

canonical coefficient of -.17) and idioms (standardized canonical coefficient of -.13). For the dependent variables, the canonical correlation was nearly equally influenced by both verbal abstract reasoning (standardized canonical coefficient of -.61) and abstract reasoning (standardized canonical coefficient of -.60).

Table 22.
Canonical Correlation between Figurative Language and Abstract Reasoning

Dimension	Canonical Correlation	*F*	df1	df2	*p*
1	.63	9.36	6	190	.00
2	.11	.58	2	96	.56

Table 23.
Standardized Canonical Coefficients of Figurative Language and Abstract Reasoning

	Dimension
Variables	1
Figurative Language	
Idioms	-.13
Proverbs	-.88
Metaphors	-.17
Abstract Reasoning	
Similarities	-.61
SHIPLEY-2 Abstraction	-.60

Hypothesis Two. Hypothesis (2) predicted that performances on the idiom, proverb and metaphor subscales would significantly contribute to the prediction of abstract reasoning abilities. To address this hypothesis two forced entry linear regressions were conducted. One regression utilized verbal abstract reasoning (as measured by Similarities) as the criterion variable. The Idiom Subscale, the Proverb Subscale, and the Metaphor Subscale were entered into the regression, along with age, total years of education, and acculturation as each of these

variables demonstrated a significant association with verbal abstract reasoning in the univariate comparisons. The overall regression was statistically significant ($R^2 = .33$, $F(6, 93) = 7.75$, $p < .001$). It was found that proverbs significantly predicted verbal abstract reasoning ($\beta = .39$, $p < .001$), while acculturation was a marginally significant predictor ($\beta = .39$, $p = .06$). The Idioms Subscale, the Metaphors Subscale, age, and total years of education did not significantly predict verbal abstract reasoning. The results from this regression are displayed in Table 24, Table 25, and Table 26 below.

Table 24.
Regression Model Summary with Verbal Abstract Reasoning as Predictor Variable

Model	*R*	R^2	Adjusted R^2	Std. Error of the Estimate
1	.58	.33	.29	3.23

Table 25.
ANOVA with Verbal Abstract Reasoning as Predictor Variable

Model		Sum of Squares	*df*	Mean Square	*F*	Sig.
1	Regression	484.65	6	80.77	7.75	.00
	Residual	969.06	93	10.42		
	Total	1453.71	99			

Table 26.
Regression Coefficients with Verbal Abstract Reasoning as Predictor Variable

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients β	*t*	Sig.
1	(Constant)	13.02	2.73		4.78	.00
	Age	.01	.05	.03	.28	.78
	Education	.26	.21	.13	1.24	.22
	Acculturation	.47	.25	.17	1.88	.06
	Idioms	-.02	.06	-.02	-.26	.79
	Proverbs	.21	.05	.39	3.88	.00
	Metaphors	.08	.06	.11	1.21	.23

Note: Acculturation was measured via the Demographic Index of Cultural Exposure.

The second regression utilized abstract reasoning (as measured by SHIPLEY-2 Abstraction) as the criterion variable. The Idiom Subscale, the Proverb Subscale, and the Metaphor Subscale were entered into the regression, along with age, total years of education, and acculturation, as each of these variables demonstrated a significant association with abstract reasoning in the univariate comparisons. The overall regression was statistically significant (R^2 = .31, $F(6, 93) = 6.82$, $p < .001$). It was found that proverbs significantly predicted abstract reasoning ($\beta = .42$, $p < .001$). The Idioms Subscale, the Metaphors Subscale, age, total years of education, and acculturation did not significantly predict abstract reasoning. The data from this regression is displayed in Table 27, Table 28, and Table 29 below.

Table 27.
Model Summary with Abstract Reasoning as Predictor Variable

Model	*R*	R^2	Adjusted R^2	Std. Error of the Estimate
1	.55	.31	.26	2.79

Table 28.
ANOVA with Abstract Reasoning as Predictor Variable

Model		Sum of Squares	*df*	Mean Square	*F*	Sig.
1	Regression	318.26	6	53.04	6.82	.00
	Residual	723.74	93	7.78		
	Total	1042.00	99			

Table 29.
Regression Coefficients with Abstract Reasoning as Predictor Variable

	Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients *β*	*t*	Sig.
1	(Constant)	8.40	2.36		3.57	.00
	Age	.05	.04	.13	1.18	.24
	Education	-.04	.18	-.02	-.22	.82
	Acculturation	.28	.22	.12	1.31	.19
	Idioms	.04	.05	.07	.76	.45
	Proverbs	.19	.05	.42	4.17	.00
	Metaphors	.02	.05	.04	.38	.71

Note: Acculturation was measured via the Demographic Index of Cultural Exposure.

Hypothesis Three. Hypothesis (3) was that the idioms, proverbs, metaphors, and the figurative language total score would each demonstrate a positive association with a) intelligence scores, b) inhibition scores, c) working memory scores, and d) cognitive flexibility scores. To address hypothesis (3) a series of Pearson's *r* correlations were conducted between the Idioms Subscale, the Proverbs Subscale, the Metaphors Subscale, Figurative Language Total Score, and scores on measures of cognition. These correlations are presented in Table 30 below.

Table 30.

Correlations Cognitive Testing and Figurative Language Scores

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Working Memory												
2. Inhibition (Time)	-.22*											
3. Inhibition (Errors)	-.16	.30**										
4. Cognitive Flexibility (Total)	-.08	-.08	.02									
5. Cognitive Flexibility (Perseverative)	.07	.03	.01	-.96**								
6. Vocabulary	-.03	.06	-.08	.14	-.13							
7. Abstract Reasoning	.18	-.10	-.10	.29**	-.26**	.40**						
8. IQ	.08	-.02	-.11	.24*	-.23*	.87**	.80**					
9. Verbal Abstract Reasoning	.17	-.04	-.14	.13	-.16	.64**	.38**	.63**				
10. Idioms	-.03	.11	-.07	-.02	.04	.24*	.25*	.29**	.16			
11. Proverbs	.04	-.08	-.04	.14	-.11	.53**	.51*	.62**	.52**	.28**		
12. Metaphors	.00	.02	-.08	.03	-.04	.40**	.25*	.39**	.32**	.08	.41**	
13. Figurative Language	.01	-.03	-.09	.11	-.09	.58**	.49**	.65**	.51**	.33**	.87**	.76**

Note: Working Memory was measured via Digit Span Backwards; Inhibition was measured via Stroop Color-Word Test Condition 3 Time and Condition 3 Errors; Cognitive Flexibility was measured via Wisconsin Card Sorting Test Total Correct and Perseverative Errors; Vocabulary was measured via SHIPLEY-2 Vocabulary; Abstract Reasoning was measured via SHIPLEY-2 Abstraction; IQ was measured via SHIPLEY-2 Composite Score; and Verbal Abstract Reasoning was measured via Similarities.

**Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed).

IQ was significantly positively associated with the Idioms Subscale, Proverbs Subscale, Metaphors Subscale, and the Figurative Language Interpretation Task total score. Inhibition (time) was not significantly associated with idioms (r = .11, p = .88), proverbs (r = -.08, p =.42), metaphors (r = .02, p = .88), or figurative language total (r = -.03, p = .77). Inhibition (errors) was not significantly associated with idioms (r = -.07, p = .50), proverbs (r = -.04, p = .73), metaphors (r = -.08, p = .41), or figurative language total (r = -.09, p = .37). Working memory was not significantly associated with idioms (r = -.03, p = .76), proverbs (r = .04, p = .69), metaphors (r = .00, p = .99), or figurative language total (r = .01, p = .89). Cognitive flexibility (total) was not significantly associated with idioms (r = -.02, p = .88), proverbs (r = .14, p = .18), metaphors (r = .03, p = .74), or figurative language total (r = .11, p = .26). Cognitive flexibility (perseverative) was not significantly associated with idioms (r = .04, p = .73), proverbs (r = -.11, p = .29), metaphors (r = -.04, p = .68), or figurative language total (r = -.09, p = .35).

Hypothesis Four. Hypothesis (4) was that male participants and female participants would vary in their figurative language interpretation abilities. To examine any potential sex differences, male participants' and female participants' scores on the figurative language subscales were examined using a series of one-way between-groups ANOVAs. No statistically significant differences were found between male participants' and female participants' scores on the Idioms Subscale (F(1, 98) = 0.00, p = .99), the Proverbs Subscale (F(1, 98) = 1.80, p = .18), the Metaphors Subscale (F(1, 98) = 2.22, p = .14), or the Figurative Language Total Score (F(1, 98) = 0.19, p = .66). On the Idioms Subscale the mean score obtained by male participants (M = 21.16) and female participants (M = 21.18) was nearly identical. On the Proverbs Subscale female participants' mean score (M = 18.90) was slightly lower than that of male participants (M = 20.94), while on the Metaphor Subscale male participants' mean score (M = 13.41) was

slightly lower than that of female participants (M = 15.25). When these scores were combined into a total figurative language interpretation score, the mean scores were nearly identical between male participants (M = 55.50) and female participants (M = 54.32).

Discussion

Purpose of the Present Study

The purpose of this research was to explore the relationship between figurative language interpretation and other cognitive abilities. Previous research has demonstrated that various forms of figurative language interpretation are uniquely affected by different lesions, clinical syndromes, and cognitive abilities (e.g., Kempler et al., 1988; Papagno et al., 2003; Papagno et al., 2006; Titone et al., 2002; Van Lancker & Kempler, 1987; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Leyhe et al., 2011; Cardoso et al., 2014; Murphy et al., 2013; McDonald et al., 2007). Despite evidence suggesting that figurative language interpretation is an important consideration in clinical practice, figurative language assessments are 1) largely unavailable for use with adult populations in psychological treatment or assessment settings, and 2) what is currently available is not popularly incorporated into clinical practice (Arcara & Bambini, 2016).

Three types of figurative language were utilized for the present research—idioms, proverbs, and metaphors—as each of these have demonstrated a significant relationship with abstract reasoning abilities in previous studies (Kempler et al., 1988; Papagno et al., 2003; Rassiga et al., 2009; Papagno et al., 2006; Papagno & Caporali, 2007; Van Lancker & Kempler, 1987; Kempler et al., 1999; Papagno & Vallar, 2001; Titone et al., 2002; Papagno et al., 1995; Amanzio et al., 2008; Monetta & Pell, 2007; Winner & Gardner, 1977; Brownell et al., 1990; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Jones & Stone, 1989; Lee & Kamhi, 1990;

Nippold & Fey, 1983; Iakimova et al., 2006; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007). Despite the evidence demonstrating a link between idioms, proverbs, metaphors and abstract reasoning abilities, no known studies had examined whether proverbs, idioms, and metaphors are equally good measures of abstract reasoning, or whether one or two types of figurative language interpretation are superior measures of abstract reasoning. Additionally, there is mixed and inconsistent evidence regarding the relationship between different types of figurative language and other cognitive abilities, such as general intelligence, inhibition, cognitive flexibility, and working memory (Brune & Bodenstein, 2005; Spoheim et al., 2003; Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008; Monetta & Pell, 2007; Iskandar, 2014).

The present research aimed to fill these gaps in the literature through two related studies. The first study was designed to create a psychometrically-sound Figurative Language Interpretation Task, which included a range of idioms, proverbs, and metaphors. Subsequently, study two assessed how interpretation of idioms, proverbs, and metaphors relate to abstract reasoning, and compared the strength of the relationship between these three types of figurative language interpretation and abstract reasoning. Additionally, study two explored the relationship between different types of figurative language interpretation and other cognitive abilities using existing neuropsychological tests that assess general intelligence, inhibition, cognitive flexibility, and working memory.

Study One

Study one assessed cognitively-healthy young adults on 60 different figurative language interpretation items. Participants' responses to each item were scored based on both accuracy and level of abstraction by two independent raters. The Figurative Language Interpretation Task pilot

stimuli included five subscale scores—one for familiar idioms, one for unfamiliar idioms, one for proverbs, and one for metaphors, as well as a total score. The specific aims of study one were 1) to ensure interrater reliability among co-raters on the Figurative Language Interpretation Task scoring system, 2) to reduce the number of items of the Figurative Language Interpretation Task, and 3) to obtain acceptable alpha levels for each subscale.

These aims were accomplished through study one. Specifically, the two independent raters who scored each item on the Figurative Language Interpretation Task were found to demonstrate good interrater reliability for the scoring system (ICC = .82, p < .001). This scoring criteria was subsequently modified and expanded to include additional example responses for 4-point, 3-point, and 2-point responses to further delineate what type of responses represented each point value. Additionally, an item discrimination analysis was utilized to remove items that did not demonstrate strong association with that item's respective subscale (r ranged from .27 to .66), and an item difficulty index was calculated to ensure the Figurative Language Interpretation Task represented a range of difficulty (M ranging from 1.59 to 3.82). Acceptable alpha levels were obtained for each of the subscales (ranging from .67 to .73) as well as for the Figurative Language Interpretation Task overall (α = .83) after using a series of deletions to items that lowered the internal consistency of each scale. The resulting Figurative Language Interpretation Task consisted of 22 items and four subscales—the Idioms Subscale, the Proverbs Subscale, the Metaphors Subscale, and the Figurative Language Interpretation Task total score. The Figurative Language Interpretation Task was used as the primary instrument of interest in study two.

Study Two

Study two utilized the Figurative Language Interpretation Task from study one, further examined its structure and described its psychometric properties, and tested four experimental

hypotheses regarding the relationship between figurative language interpretation and cognitive abilities. Specifically, study two assessed how interpretation of idioms, proverbs, and metaphors relate to abstract reasoning, and examined the strength of the relationship between these three types of figurative language interpretation and abstract reasoning. Additionally, study two explored the relationship between different types of figurative language interpretation and general intelligence, inhibition, cognitive flexibility, as well as working memory.

The aims of study two were 1) to examine the final factor structure of the measure, and 2) to assess the alpha levels for any identified scales. In addressing the first aim, an exploratory factor analysis utilizing a minimum average partial test was conducted and did not reveal any discernible factor structure. As a result, no clear scales were identifiable. Although several proverbs appeared to emerge as the most useful factor, there was no clear pattern among the other factor loadings. However, Cronbach's alpha was calculated for the three figurative language subscales—the Idioms Subscale (α = .54), the Proverbs subscale (α = .65), and the Metaphors Subscale (α = .60)—and the Figurative Language Total Score (α = .76). The idioms, proverbs, and metaphors subscales did not meet the minimum internal consistency threshold of .70 recommended for exploratory measures (Hair et al., 2014; Nunnally, 1978). However, the items included in these measures were intentionally selected to represent a range of difficulty, which may have lowered the alpha levels of these scales as an assumption of Cronbach's alpha is that all items are equally representative measures of a construct.

Overall, the psychometric properties of the Figurative Language Interpretation Task revealed that the measure represented a range of difficulty, each item correlated significantly with its respective subscale and with the Figurative Language Interpretation Task total score, and the Figurative Language Interpretation Task total score demonstrated acceptable internal

consistency (α = .76; Hair et al., 2014; Nunnally, 1978). Furthermore, inter-rater reliability of the modified and expanded scoring system for the Figurative Language Interpretation Task was excellent (*ICC* = .92, $p < .001$), and scores on each item from study one and study two demonstrated a significant positive association with one another when compared using a Spearman Rho rank order correlation (ρ = .89, $p < .001$). Together, this evidence suggests that the Figurative Language Interpretation Task is a viable and psychometrically-sound method of assessing figurative language interpretation abilities.

Prior to engaging in our hypothesis testing, a series of univariate comparisons were conducted. Results of these analyses revealed that age demonstrated a significant positive association with idiom interpretation, proverb interpretation, figurative language interpretation (total score), vocabulary, abstract reasoning, IQ, and verbal abstract reasoning, but not with metaphor interpretation. By contrast, age was not associated with working memory, inhibition (errors), cognitive flexibility (total), or cognitive flexibility (errors). These findings align with previous findings which suggest that crystallized intelligence maintains or increases across the lifespan, whereas fluid intellectual abilities peak in young adulthood and gradually decline across the lifespan (Wang & Kaufman, 1993; Cunningham et al., 1975). Additionally, education was positively associated with idiom interpretation, proverb interpretation, figurative language interpretation (total score), vocabulary, abstract reasoning, IQ, and verbal abstract reasoning. These findings also align with our understanding of the relationship between education and intelligence. Specifically, previous literature has found significant positive associations between higher educational attainment and IQ, and more specifically, between higher educational attainment and verbal expression/comprehension (Wechsler 1958; Kaufman & Lichtenberger, 2006; Manly et al., 2000). Among personality traits, extraversion was positively associated with

proverb interpretation and figurative language (total score). No other significant associations were found between any of the figurative language subscales or any of the Big Five personality traits. One explanation for this finding between two of the figurative language scales and extraversion may be that the more socially engaged, or extraverted, an individual is, the more they are exposed to and have experience engaging with various types of figurative language.

Hypothesis Testing

Hypothesis One. Hypothesis one posited that the Figurative Language Interpretation Task scales would collectively demonstrate a positive association with other measures of abstract reasoning as existing evidence suggests that figurative language interpretation is linked to abstract reasoning (e.g., Elmore & Gorham, 1957; Delis et al., 2001b; Iskandar, 2014). A series of Pearson's *r* correlations found that idioms, proverbs, metaphors, and figurative language interpretation (total score) were all positively correlated with a measure of abstract reasoning. Additionally, proverbs, metaphors, and figurative language interpretation (total score) were all positively correlated with a measure of verbal abstract reasoning; however, idioms were not. Furthermore, results of a canonical correlation indicated that there was a significant positive association between the three figurative language types (idioms, proverbs, and metaphors) and two measures of abstract reasoning. Interestingly, proverbs demonstrated the strongest association with abstract reasoning, followed by metaphors, with idioms having the weakest association. Together, this evidence suggests that figurative language interpretation is strongly linked to abstract reasoning abilities. However, the strength of the relationship between figurative language and abstract reasoning appears to vary based on the type of figurative language used, with proverbs demonstrating the strongest association with abstract reasoning abilities.

Overall, the results of these Pearson's *r* correlations and the canonical correlation provide evidence for the utility of figurative language interpretation as a valid measure of abstract reasoning abilities. More specifically, these results also indicate that proverb interpretation is the strongest measure of abstract reasoning abilities as compared to idiom and metaphor interpretation. In current clinical practice, the most widely available assessment of figurative language interpretation is the D-KEFS Proverb Test (Delis et al., 2001a), and the results of this study support the use of this instrument as a measure of abstract reasoning.

Through a variety of analyses proverbs appeared to be the best measure of abstract reasoning abilities when compared to other types of figurative language interpretation, and there are several possible explanations for this finding. For example, existing evidence suggests that impaired metaphor and idiom interpretation is associated with deficits in abstract reasoning (Kempler et al., 1988; Papagno et al., 2003; Rassiga et al., 2009; Papagno et al., 2006; Papagno & Caporali, 2007; Van Lancker & Kempler, 1987; Kempler et al., 1999; Papagno & Vallar, 2001; Titone et al., 2002; Papagno et al., 1995; Amanzio et al., 2008; Monetta & Pell, 2007; Winner & Gardner, 1977; Brownell et al., 1990; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Jones & Stone, 1989; Lee & Kamhi, 1990; Nippold & Fey, 1983; Iakimova et al., 2006; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007), however, the results of this canonical correlation, which utilized data collected within a cognitively and psychiatrically healthy population, did not find strong relationships between idiom interpretation or metaphor interpretation and abstract reasoning abilities. One possible explanation for this finding is that the impairments found in previous studies were products of the individuals in those studies having known cognitive and (or) psychiatric impairments that may impact multiple abilities or result in more pervasive cognitive difficulties. For example, a

psychiatric disorder such as schizophrenia may result in deficits in abstract reasoning and deficits in idiom or metaphor interpretation, however, this would not necessarily imply that the deficits in idiom or metaphor interpretation were found as a direct result of abstract reasoning impairments. Instead, it is possible that abstract reasoning and idiom or metaphor interpretation were separately impacted by schizophrenia's neurobiological processes and the relationship between these variables was impacted by the participants' broader set of cognitive deficits. It may also be the case that figurative language interpretation functions differently in cognitively and psychologically healthy individuals than in clinical populations. The results of this study, which utilized a cognitively and psychologically healthy population, may provide supportive evidence for either of these conclusions. Alternatively, the variations in the relationship between idiom, metaphor, and proverb interpretation may provide supportive evidence for the conclusion that idioms, proverbs, and metaphors involve different types or levels of cognitive processing, which has been posited in previous studies (Searle, 1979; Papagno, 2010; Honeck & Hoffman, 1980; Mordkoff & Yantis, 1991; Ortony, 1978; Blasko & Connine, 1993; Giora et al., 2012).

A related finding was that idioms were not associated with the measure of verbal abstract reasoning and demonstrated the smallest contribution in the canonical correlation. Idioms are defined as expressions that convey general information using socially or culturally known definitions (Papagno, 2010; Jackendoff, 1995), and accordingly are sometimes thought to be stored in semantic memory together with word meanings and concepts (Jackendoff, 1995). Additionally, studies have shown that the figurative meaning of highly familiar or overlearned phrases is processed before the literal meaning (e.g., *"She is a warm person"* was more quickly understood as a personality trait than an actual temperature; Blasko & Connine, 1993; Giora et al., 2012), and that these phrases no longer require a mental search (i.e., higher order cognitive

abilities) to find the nonliteral definition (Iskandar, 2014). The results from this study provide support for this idea, as idioms did not correlate with verbal abstract reasoning, demonstrated the weakest overall relationship with abstract reasoning, but were significantly associated with general vocabulary knowledge, which supports the idea that idioms are stored in semantic memory.

Hypothesis Two. Hypothesis two stated that performances on the idiom, proverb, and metaphor subscales would significantly predict abstract reasoning abilities. To address this hypothesis two forced entry linear regressions were conducted. The Idiom Subscale, the Proverb Subscale, and the Metaphor Subscale were entered into the regressions, along with age, total years of education, and level of acculturation, as each of these variables demonstrated a significant association with measures of abstract reasoning in the univariate comparisons. In both regressions, the only predictor variable to reach significance was the Proverbs Subscale, indicating that proverb interpretation appears to be the strongest predictor of abstract reasoning abilities. The results of these regressions, along with the results of the aforementioned Pearson's *r* correlations and canonical correlation, further support the idea that proverb interpretation is the best predictor of abstract reasoning abilities. These results, again, provide support for the use of the D-KEFS Proverb Test (Delis et al., 2001a) in clinical practice as a measure of abstract reasoning abilities.

Similar to the findings for hypothesis one, idiom interpretation and metaphor interpretation did not demonstrate a strong relationship with abstract reasoning abilities, which was not expected based on previous research (Kempler et al., 1988; Papagno et al., 2003; Rassiga et al., 2009; Papagno et al., 2006; Papagno & Caporali, 2007; Van Lancker & Kempler, 1987; Kempler et al., 1999; Papagno & Vallar, 2001; Titone et al., 2002; Papagno et al., 1995;

Amanzio et al., 2008; Monetta & Pell, 2007; Winner & Gardner, 1977; Brownell et al., 1990; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Jones & Stone, 1989; Lee & Kamhi, 1990; Nippold & Fey, 1983; Iakimova et al., 2006; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007). As stated for hypothesis one, there may be several explanations for the lack of findings between idiom interpretation, metaphor interpretation, and abstract reasoning abilities. Specifically, impairments found in previous studies may have been products of the individuals in those studies having known cognitive and (or) psychiatric impairments, figurative language interpretation may function differently in cognitively and psychologically healthy individuals than in clinical populations, and (or) different types of figurative language may involve different types or levels of cognitive processing. Additionally, idioms may be more reflective of vocabulary knowledge than of abstract reasoning abilities, which may also partially explain why proverbs appeared to be the best measure of abstract reasoning abilities in the present study. The results from these regressions may provide supportive evidence for any of these conclusions.

It is also worth noting that the only other predictor variable that approached significance in the regression analyses was acculturation. Acculturation was included in the current research as proverbs and idioms are, by definition, culturally bound phrases (Leyhe et al., 2011; Gibbs & Beitel, 1995; Papagno, 2010; Jackendoff, 1995) that rely on cultural exposure, knowledge of social norms, and common social experience. Additionally, environmental exposure to these phrases likely plays a key role in gaining knowledge and understanding of these types of phrases. Accordingly, the results from the present study found that acculturation was positively correlated with proverb interpretation, the metaphor interpretation, and figurative language interpretation (total score). Furthermore, acculturation was positively correlated with abstract reasoning and

verbal abstract reasoning, and emerged as marginally significant in our regression analysis predicting verbal abstract reasoning performance. Together these results indicate that acculturation is likely an important covariate that influences abstract reasoning abilities. One explanation could be that this is due to language exposure in general (i.e., the more one is around the language, the better one understands that language's figurative phrases); however, a series of one-way between-groups ANOVAs did not find any statistically significant differences between native English speakers' and nonnative English speakers' interpretation of idioms, proverbs, metaphors, or figurative language (total score). However, it is also important to consider the fact that globally there appears to be an emphasis on the English language and an emphasis on American culture in general. Due to this, individuals who scored lower on the measure of acculturation may have also had a higher level of exposure to typical American culture than would be reflected in their acculturation score alone. The specific impacts of this are unknown and were beyond the scope of the present study but are nonetheless important to consider as it may have also impacted the present findings related to acculturation.

Taken together, however, our findings suggests that level of acculturation influences figurative language interpretation and abstract reasoning abilities, but that this influence cannot be accounted for by knowledge of and exposure to the English language in general. Instead, it appears that knowledge of and experience with the culture (e.g., social norms, common social experiences, etc.) may play an important role in understanding figurative language.

Hypothesis Three. Hypothesis three stated that figurative language interpretation would demonstrate a positive association with a) intelligence scores, b) inhibition scores, c) working memory scores, and d) cognitive flexibility scores. This was believed to be the case as each of these cognitive abilities has been associated with at least one type of figurative language

interpretation in previous studies (Brune & Bodenstein, 2005; Spoheim et al., 2003; Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008; Monetta & Pell, 2007; Iskandar, 2014). However, the findings have largely been inconsistent across these previous studies.

In the current study, idioms, proverbs, metaphors, and figurative language interpretation (total score) were all significantly positively associated with general intelligence. However, no associations were found between any of the figurative language interpretation scores and inhibition, working memory, or cognitive flexibility. Regarding inhibition abilities, previous cognitive processing theories (Searle, 1979; Papagno, 2010) have posited that figurative language is an effortful two-step process that likely includes an inhibitory process in order to reach the correct, abstract meaning of figurative phrases. However, at present, the most widely supported theories suggest that understanding of figurative language occurs through parallel and direct mental search channels that are not contingent on inhibitory processes (Honeck & Hoffman, 1980; Mordkoff & Yantis, 1991; Ortony, 1978; Johnson 1996; Glucksberg et al., 1982). The lack of relationship between figurative language interpretation and inhibition abilities found in our results may provide further support for the parallel and direct cognitive processing theory. Previous studies (Brune & Bodenstein, 2005; Spoheim et al., 2003; Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008; Monetta & Pell, 2007; Iskandar, 2014) have indicated that working memory is associated with idiom and metaphor interpretation, whereas cognitive flexibility is associated with proverb and idiom interpretation. However, it is also important to note that the findings in previous studies have been inconsistent and were not able to be replicated in the present study. As mentioned with hypothesis one and two, the impairments found in previous studies may have been products of the individuals in those studies having known cognitive and (or) psychiatric impairments that globally impact cognition. Additionally,

the previous findings may have resulted from variations in methodology and analyses used across these studies (e.g., through the use of different figurative language or cognitive measures).

In summary, this study resulted in a lack of significant findings among figurative language interpretation and inhibition, working memory, and cognitive flexibility, while identifying a strong relationship between figurative language interpretation and abstract reasoning abilities. Together this data supports the conclusion that figurative language interpretation is a valid measure of abstract reasoning that does not appear to be impacted or influenced by a heterogeneous mix of cognitive abilities (i.e., inhibition, working memory, or cognitive flexibility) within a healthy adult population.

Hypothesis Four. Hypothesis four predicted that male participants and female participants would vary in their figurative language interpretation abilities, with female participants scoring higher on average, as previous research has indicated that the ratio of language impairment between females and males is 2:5 (Ellis Weismer et al., 2021). A series of one-way between-groups ANOVAs did not reveal any statistically significant differences between male participants' and female participants' scores on idioms, proverbs, metaphors, or figurative language (total score). On idiom interpretation the mean score obtained by male and female participants was nearly identical, on proverb interpretation female participants' mean score was slightly lower than that of male participants, and on metaphor interpretation male participants' mean score was slightly lower than that of female participants. When these scores were combined into a total figurative language interpretation score, the mean scores were nearly identical between male participants and female participants. The lack of differences found in our data are likely a result of the inclusion and exclusion criteria used for this study. Although there is a two female to five male ratio of language impairment, our participants only included

individuals who were known to be cognitively and psychiatrically healthy. This effectively eliminated any individual with known language impairment, and in general, we do not expect cognitively healthy adult male participants and female participants to demonstrate differences in cognitive abilities (Hyde & Linn, 1988; Hyde et al., 1990; Hyde, 2005; Spelke, 2005). Instead, the findings from this study support the conclusion that cognitively and psychiatrically healthy individuals do not demonstrate sex differences on tests of figurative language interpretation.

Limitations and Future Directions

There are several limitations to the present studies. Regarding the participant population, both studies utilized a cognitively and psychiatrically healthy population, and as a result, our data may not have been sensitive to some of the impairments in cognition and figurative language interpretation which had been found in previous studies. In addition, our population primarily consisted of students at a university in the midwestern region of the country, which may limit the generalizability of the results. Relatedly, the findings from these studies suggest that acculturation is an important covariate to consider when assessing figurative language interpretation, however, understanding the specific impact of acculturation was beyond the scope of the present study. As such, the degree to which level of acculturation influenced our findings is unknown and should be considered a potential limitation of the present findings.

Another potential limitation to the present study is that corrections were not made for potential familywise errors. However, in exploratory research, such as the present study, it is recommended that familywise error is conceptualized in terms of the number of analyses used for each hypothesis (Rubin, 2017). In the present study, the number of analyses employed per hypothesis was low and the associated p-values of significant findings were largely above the 0.01 threshold of significance.

Additionally, the Figurative Language Interpretation Task was designed to only assess expressive abilities (not receptive abilities), which may be more heavily influenced by verbal expression abilities and general knowledge of vocabulary (ASHA, n.d.d.; ASHA, n.d.e.). For example, when presented with the phrase *"Too many cooks spoil the broth,"* a participant may understand the phrase, and even know a correct social response, but they may be limited in their ability to articulate what that phrase means in their own words. In this way, our focus on expressive language abilities may have limited the scope with which we were able to define and examine figurative language interpretation. Furthermore, when selecting the items for the Figurative Language Interpretation Task, level of abstractness and level of exposure to the phrases were considered; however, it is possible that a different set of metaphors, proverbs, and idioms may have elicited slightly different results (for other examples, see: Benjafield et al., 1993; Billow, 1975; Katz et al., 1988; Campbell & Raney, 2016; Titone & Connine, 1994; Libben & Titone, 2008; Tanner & Bulkes, 2017). Previous research has indicated that figurative language comes in high varied forms (Nunberg et al., 1994; Cacciari & Glucksberg, 1995; Gibbs & Colston, 2006; Papagno & Genoni, 2004), and various factors influence the manner in which figurative language is cognitively processed, with characteristics such as context, degree of complexity, and degree of familiarity influencing which type of processing occurs (Blasko & Connine, 1993; Giora et al., 2012; Ortony, 1978). Some phrases automatically elicit figurative definitions, some phrases automatically elicit literal definitions, and some phrases equally elicit figurative and literal definitions (Blasko & Connine, 1993; Giora et al., 2012; Ortony, 1978; Searle, 1979; Papagno, 2010; Honeck & Hoffman, 1980). Additionally, some phrases may have been more concrete and may not have required a high level of abstraction, for example, the metaphor *"genes are blueprints,"* may have elicited more concrete answers than a metaphor

such as *"the mind is a sponge."* As a result, the items chosen could have influenced the degree to which participants felt the need to provide abstract responses.

Given these limitations, there are several future directions to expand on the current findings. Specifically, future studies could attempt to replicate these findings using the newly-created Figurative Language Interpretation Task in a wider variety of samples, both demographically (e.g., non-student samples, various geographic regions) and among different clinical populations. Specifically, the Figurative Language Interpretation Task could be utilized to examine any clinical correlates with neurocognitive or psychiatric conditions as these types of samples may be more sensitive to differences in cognition that were not found in the current sample (as was the case in previous studies; e.g., Kempler et al., 1988; Papagno et al., 2003; Papagno et al., 2006; Titone et al., 2002; Van Lancker & Kempler, 1987; Cutting & Murphy, 1990; Iakimova et al., 2006; Drury et al., 1998; Kircher et al., 2007; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Leyhe et al., 2011; Cardoso et al., 2014; Murphy et al., 2013; McDonald et al., 2007). Relatedly, in this study, figurative language interpretation appears to be related to acculturation. As such, future studies could further examine the relationship between figurative language interpretation and acculturation, and more specifically, could examine its utility in a sample of less acculturated individuals.

Furthermore, modified versions of the Figurative Language Interpretation Task could be created to expand and test a different set of metaphors, proverbs, and idioms to determine whether our results replicate with a different set of phrases. Specifically, versions of the Figurative Language Interpretation Task could be created to assess not only expressive language abilities, but also receptive language abilities and socially appropriate responses. To assess receptive language abilities, a multiple-choice format of the items used in the Figurative

Language Interpretation Task could be utilized to identify whether participants can accurately identify correct responses, regardless of their ability to verbally express their own definitions of the phrases. This is commonly utilized in other measures of figurative language interpretation (Delis et al., 2001a; Iskandar & Baird, 2014). Relatedly, another version of the study could be designed in which participants are asked to give an appropriate social response to each phrase which could be structured to include vignettes that provide the participant with a brief description of a social interaction or conversation with includes the use of a figurative language statement. The participant would then be prompted to provide a response indicating how they might respond to this social interaction or proceed in the conversation. The information gathered from this type of task may aid researchers and clinicians in gathering information about one's ability to socially use and engage with figurative phrases regardless of one's ability to explain or define the phrase. As other versions of the Figurative Language Interpretation Task are developed and tested, it would also be interesting to examine the ecological validity of this measure. Specifically, it could be useful to examine how scores on the Figurative Language Interpretation Task relate to self-report measures of behavioral and psychosocial functioning in daily life. Given that figurative language interpretation is a common skill required in communicative exchanges, it may be the case that assessment of figurative language is a more ecologically valid method of assessing deficits in daily functioning when compared to traditional neurocognitive testing.

Conclusions

In conclusion, this was a two-part study that intended to explore the relationship between figurative language interpretation, abstract reasoning, and other cognitive abilities. To do this, three types of figurative language were examined—idioms, proverbs, and metaphors—as each of these have demonstrated a significant relationship with abstract reasoning abilities in previous studies (e.g., Papagno et al., 2006; Papagno & Caporali, 2007; Van Lancker & Kempler, 1987; Kempler et al., 1999; Papagno & Vallar, 2001; Titone et al., 2002; Amanzio et al., 2008; Monetta & Pell, 2007; Happe, 1993; Gold & Faust, 2010; Giora et al., 2012; Iakimova et al., 2006; Cutting & Murphy, 1990; Drury et al., 1998; Kircher et al., 2007). Additionally, there is evidence to suggest that figurative language interpretation, as represented by idioms, proverbs, and metaphors, is related to general intelligence, inhibition, cognitive flexibility, and working memory (Brune & Bodenstein, 2005; Spoheim et al., 2003; Papagno & Caporali, 2007; Schettino et al., 2010; Amanzio et al., 2008; Monetta & Pell, 2007; Iskandar, 2014).

Study one assessed cognitively-healthy young adults on 60 different figurative language interpretation items, which included a range of idioms, proverbs, and metaphors. This study used these items to create a novel 22-item Figurative Language Interpretation Task, with good interrater reliability for its scoring system and acceptable alpha levels for all four subscales of the measure. Study two utilized the Figurative Language Interpretation Task from study one and further examined its structure and described its psychometric properties. Overall, the psychometric properties of the Figurative Language Interpretation Task represented a range of difficulty; each item correlated significantly with its respective subscale and with the Figurative Language Interpretation Task total score, and the Figurative Language Interpretation Task demonstrated acceptable internal consistency (Hair et al., 2014; Nunnally, 1978). Furthermore,

inter-rater reliability of the modified and expanded scoring system for the Figurative Language Interpretation Task was excellent and scores on each item from study one and study two demonstrated a significant positive association with one another when compared using a Spearman Rho rank order correlation.

Additionally, study two tested four experimental hypotheses to assess the relationship between figurative language interpretation, as measured by the Figurative Language Interpretation Task, abstract reasoning, and other cognitive abilities. Results from Pearson's *r* correlations, a canonical correlation, and a series of regressions indicated that figurative language interpretation is strongly linked to abstract reasoning abilities. However, the strength of the relationship between figurative language and abstract reasoning appears to vary based on the type of figurative language used. More specifically, proverbs appeared to demonstrate the strongest correlation with abstract reasoning abilities, emerged as the strongest predictor of abstract reasoning abilities, and was the only predictor variable to reach statistical significance in both regressions. Together, this suggests that figurative language interpretation, and specifically proverb interpretation, is a valid method for assessing abstract reasoning abilities in a cognitively and psychologically healthy population. Additionally, this data supports the use of the newly created Figurative Language Interpretation Task as a valid measure in the assessment of abstract reasoning abilities.

Furthermore, the only other cognitive ability that demonstrated any relationship with figurative language interpretation was general intelligence, which demonstrated a strong positive association with the Figurative Language Interpretation Task. This study had a lack of significant findings among language interpretation and inhibition, working memory, and cognitive flexibility. These results, in combination with the significant findings among figurative language

interpretation and abstract reasoning may support the conclusion that figurative language interpretation, and specifically proverb interpretation, is a valid method of assessing abstract reasoning that does not appear to be impacted or influenced by a heterogeneous mix of various, unrelated cognitive abilities (i.e., inhibition, working memory, or cognitive flexibility). Future studies may continue to expand this work into a wider variety of sample populations and through the use of modified and expanded versions of the Figurative Language Interpretation Task.

REFERENCES

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433-459. https://doi.org/10.1002/wics.101

Ahrens, K., Liu, H. L., Lee, C. Y., Gong, S. P., Fang, S. Y., & Hsu, Y. Y. (2007). Functional MRI of conventional and anomalous metaphors in Mandarin Chinese. *Brain and Language, 100*(2), 163-171. https://doi.org/10.1016/j.bandl.2005.10.004

Amanzio, M., Geminiani, G., Leotta, D., & Cappa, S. (2008). Metaphor comprehension in Alzheimer's disease: Novelty matters. *Brain and Language, 107*(1), 1-10. https://doi.org/10.1016/j.bandl.2007.08.003

American Psychiatric Association, (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5.* Arlington, VA. https://doi.org/10.1176/appi.books.9780890425596

American Psychological Association. (2017). *Multicultural Guidelines: An Ecological Approach to Context, Identity, and Intersectionality*. American Psychological Association.

American Psychological Association. (2019*). APA Guidelines on Race and Ethnicity in Psychology.* American Psychological Association.

American Speech-Language-Hearing Association. (n.d.a). *Components of Social Communication* (Practice Portal). American Speech-Language-Hearing Association. Retrieved July 1, 2022, from https://www.asha.org/practice-portal/clinical-topics/social-communication-disorder/components-of-social-communication/

American Speech-Language-Hearing Association. (n.d.b). *Social Communication Disorder* (Practice Portal). American Speech-Language-Hearing Association. Retrieved July 1, 2022, from https://www.asha.org/practice-portal/clinical-topics/social-communication-disorder/

American Speech-Language-Hearing Association. (n.d.c). *Social Communication* (Practice Portal). American Speech-Language-Hearing Association. Retrieved July 1, 2022, from https://www.asha.org/public/speech/development/social-communication/

American Speech-Language-Hearing Association. (n.d.d). *What is speech? What is language?* American Speech-Language-Hearing Association. Retrieved April 21, 2023, from https://www.asha.org/public/speech/development/speech-and-language/

American Speech-Language-Hearing Association. (n.d.e). *Language in Brief.* American Speech-Language-Hearing Association. Retrieved April 21, 2023, from https://www.asha.org/practice-portal/clinical-topics/spoken-language-disorders/language-in-brief/

Arcara, G., & Bambini, V. (2016). A test for the assessment of pragmatic abilities and cognitive substrates (APACS): Normative data and psychometric properties. *Frontiers in Psychology, 7*, 70. https://doi.org/10.3389/fpsyg.2016.00070

Ardila, A., Pineda, D., & Rosseli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 15*, 31–36. https://doi.org/10.1093/arclin/15.1.31

Armbruster, D. J., Ueltzhöffer, K., Basten, U., & Fiebach, C. J. (2012). Prefrontal cortical mechanisms underlying individual differences in cognitive flexibility and stability. *Journal of Cognitive Neuroscience, 24*(12), 2385-2399. https://doi.org/10.1162/jocn_a_00286

Aussie Slang Dictionary. Australia Day. (2022). Retrieved June 22, 2022, from https://www.australiaday.com.au/fun-activities/browse-the-aussie-slang-dictionary/letter/l/

Ayto, J., & Oxford University Press. (2020). *Oxford Dictionary of Idioms (Fourth Edition).* Oxford University Press. https://doi.org/10.1093/acref/9780198845621.001.0001

Baker, J. (2013). Key Components of Social Skills Training. *Teaching Social Skills to People with Autism: Best Practices in Individualizing Interventions.* Bethesda, MD: Woodbine House, Inc.

Balconi, M. (2010). Intentions and communication: Cognitive strategies, metacognition and social cognition. In *Neuropsychology of Communication* (pp. 159-175). Springer, Milano. https://doi.org/10.1007/978-88-470-1584-5_9

Balconi, M., & Amenta, S. (2010). From Pragmatics to Neuropragmatics. In *Neuropsychology of Communication* (pp. 93-109). Springer, Milano. https://doi.org/10.1007/978-88-470-1584-5_5

Bambini V. (2010). Neuropragmatics: a foreword. *Italian Journal of Linguistics.* 22, 1–20.

Bambini V., & Bara B. G. (2012). Neuropragmatics, in *Handbook of Pragmatics*, (Eds.) Östman J.-O., Verschueren J. (Amsterdam: John Benjamins), 1–21. https://doi.org/10.1075/hop.16.neu2

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist, 13*, 283–292. https://doi.org/10.1076/clin.13.3.283.1743

Benjafield, J., Frommhold, K., Keenan, T., Muckenheim, R., & Mueller, D. (1993). Imagery, concreteness, goodness, and familiarity ratings for 500 proverbs sampled from the Oxford Dictionary of English Proverbs. *Behavior Research Methods, Instruments, & Computers, 25*(1), 27-40. https://doi.org/10.3758/BF03204446

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology, 39,* 15–22. https://doi.org/10.1080/00221309.1948.9918159

Billow, Richard M. (1975). A cognitive developmental study of metaphor comprehension. *Developmental Psychology, 11*(4), 415-423. https://doi.org/10.1037/h0076668

Bishop, D. (2006). *Children's Communication Checklist (CCC-2).*

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 295-308. https://doi.org/10.1037/0278-7393.19.2.295

Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012). Old proverbs in new skins–an fMRI study on defamiliarization. *Frontiers in Psychology, 3,* 204. https://doi.org/10.3389/fpsyg.2012.00204

Bondi, M. W., Serody, A. B., Chan, A. S., Eberson-Schumate, S. C., Delis, D. C., Hansen, L. A., & Salmon, D. P. (2002). Cognitive and neuropathologic correlates of Stroop Color-Word Test performance in Alzheimer's disease. *Neuropsychology, 16,* 335–343. https://doi.org/10.1037/0894-4105.16.3.335

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., & Frith, D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language A positron emission tomography activation study. *Brain, 117*(6), 1241-1253. https://doi.org/10.1093/brain/117.6.1241

Bottini, G., Paulesu, E., Sterzi, R., Warburton, E., Wise, R. J. S., Vallar, G., & Frith, C. D. (1995). Modulation of conscious experience by peripheral sensory stimuli. *Nature, 376*(6543), 778-781. https://doi.org/10.1038/376778a0

Brownell, H. H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients. *Brain and Language, 18*(1), 20-27. https://doi.org/10.1016/0093-934X(83)90002-0

Brownell, H. H., Simpson, T. L., Bihrle, A. M., Potter, H. H., & Gardner, H. (1990). Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. *Neuropsychologia, 28*(4), 375-383. https://doi.org/10.1016/0028-3932(90)90063-T

Brownell, H. H., Simpson, T. L., Bihrle, A. M., Potter, H. H., & Gardner, H. (1990). Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. *Neuropsychologia, 28*(4), 375-383. https://doi.org/10.1016/0028-3932(90)90063-T

Brüne, M., & Bodenstein, L. (2005). Proverb comprehension reconsidered— 'theory of mind' and the pragmatic use of language in schizophrenia. *Schizophrenia Research*, *75*(2-3), 233-239. https://doi.org/10.1016/j.schres.2004.11.006

Bryan, K. L. (1995). *The Right Hemisphere Language Battery, 2nd Edition*. London: Whurr Publisher.

Bulkes, N. Z., & Tanner, D. (2017). "Going to town": Large-scale norming and statistical analysis of 870 American English idioms. *Behavior Research Methods*, 49(2), 772-783. https://doi.org/10.3758/s13428-016-0747-8

Burgess, C., & Chiarello, C. (1996). Neurocognitive mechanisms underlying metaphor comprehension and other figurative language. *Metaphor and Symbol*, *11*(1), 67-84. https://doi.org/10.1207/s15327868ms1101_4

Cacciari, C., & Glucksberg, S. (1995). Understanding idioms: Do visual images reflect figurative meanings? *European Journal of Cognitive Psychology*, *7*(3), 283-305. https://doi.org/10.1080/09541449508402450

Campbell, S. J., & Raney, G. E. (2016). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behavior Research Methods*, *48*(1), 330-340. https://doi.org/10.3758/s13428-015-0575-2

Canal, P., Vespignani, F., Molinaro, N., & Cacciari, C. (2010). Anticipatory mechanisms in idiom comprehension: Psycholinguistic and electrophysiological evidence. In *Neuropsychology of Communication* (pp. 131-144). Springer, Milano. https://doi.org/10.1007/978-88-470-1584-5_7

Cardoso, S., Silva, D., Maroco, J., De Mendonça, A., & Guerreiro, M. (2014). Non-literal language deficits in mild cognitive impairment. *Psychogeriatrics*, *14*(4), 222-228. https://doi.org/10.1111/psyg.12101

Caron, P. O. (2019). Minimum average partial correlation and parallel analysis: The influence of oblique structures. *Communications in Statistics-Simulation and Computation*, *48*(7), 2110-2117. https://doi.org/10.1080/03610918.2018.1433843

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1. https://doi.org/10.1037/h0046743

Champagne-Lavau, M., Stip, E., & Joanette, Y. (2007). Language functions in right-hemisphere damage and schizophrenia: apparently similar pragmatic deficits may hide profound differences. *Brain*, *130*(2), e67-e67. https://doi.org/10.1093/brain/awl311

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386*(6625), 604-608. https://doi.org/10.1038/386604a0

Collins, A., Lockton, E., & Adams, C. (2014). Metapragmatic explicitation ability in children with typical language development: Development and validation of a novel clinical assessment. *Journal of Communication Disorders, 52,* 31-43. https://doi.org/10.1016/j.jcomdis.2014.07.001

Coulson, S., & Van Petten, C. (2002). *Conceptual integration and metaphor: An event- related potential study.* Springer New York. https://doi.org/10.3758/BF03195780

Cruz, R. A., Wilkinson, A. V., Bondy, M. L., & Koehly, L. M. (2012). Psychometric evaluation of the Demographic Index of Cultural Exposure (DICE) in two Mexican-origin community samples. *Hispanic Journal of Behavioral Sciences, 34*(3), 404-420. https://doi.org/10.1177/0739986312449426

Cuellar, I., Arnold, B., & Gonzalez, G. (1995). Cognitive referents of acculturation: Assessment of cultural constructs in Mexican Americans. *Journal of Community Psychology, 23*(4), 339-356. https://doi.org/10.1002/1520-6629(199510)23:4<339:AID-JCOP2290230406>3.0.CO;2-7

Cunningham, W. R., Clayton, V., & Overton, W. (1975). Fluid and crystallized intelligence in young adulthood and old age. *Journal of Gerontology*, 30(1), 53-55. https://doi.org/10.1093/geronj/30.1.53

Cutting, J., & Murphy, D. (1990). Impaired ability of schizophrenics, relative to manics or depressives, to appreciate social knowledge about their culture. *The British Journal of Psychiatry, 157*(3), 355-358. https://doi.org/10.1192/bjp.157.3.355

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001a). Delis-Kaplan Executive Function System. https://doi.org/10.1037/t15082-000

Delis, D. C., Kaplan, E., & Kramer, J., (2001b). *Delis-Kaplan Executive Function System (D-KEFS) Technical Manual.* San Antonio, TX: The Psychological Corporation. https://doi.org/10.1037/t15082-000

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135. https://doi.org/10.1146/annurev-psych-113011-143750

Drury, V. M., Robinson, E. J., & Birchwood, M. (1998). 'Theory of mind' skills during an acute episode of psychosis and following recovery. *Psychological Medicine, 28*(5), 1101-1112. https://doi.org/10.1017/S0033291798006850

Eaves, L. C., & Ho, H. H. (2008). Young adult outcome of autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(4), 739-747. https://doi.org/10.1007/s10803-007-0441-x

Ellis Weismer, S., Rubenstein, E., Wiggins, L., & Durkin, M. S. (2021). A preliminary epidemiologic study of social (pragmatic) communication disorder relative to autism spectrum disorder and developmental disability without social communication deficits. *Journal of Autism and Developmental Disorders, 51*(8), 2686–2696. https://doi.org/10.1007/s10803-020-04737-4

Elmore, C. M., & Gorham, D. R. (1957). Measuring the impairment of the abstracting function with the Proverbs Test. *Journal of Clinical Psychology*. https://doi.org/10.1002/1097-4679(195707)13:3<263:AID-JCLP2270130308>3.0.CO;2-C

Eviatar, Z., & Just, M. A. (2006). Brain correlates of discourse processing: An fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia, 44*(12), 2348-2359. https://doi.org/10.1016/j.neuropsychologia.2006.05.007

Fellbaum, C. (2007). Idioms and collocations: From corpus to electronic lexical resource. *Continuum, Birmingham.*

Ferland, M. B., Ramsay, J., Engeland, C., & O'Hara, P. (1998). Comparison of the performance of normal individuals and survivors of traumatic brain injury on repeat administrations of the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology, 20*, 473–482. https://doi.org/10.1076/jcen.20.4.473.1475

Fogliata, A., Rizzo, S., Reati, F., Miniussi, C., Oliveri, M., & Papagno, C. (2007). The time course of idiom processing. *Neuropsychologia*, *45*(14), 3215-3222. https://doi.org/10.1016/j.neuropsychologia.2007.06.009

Gardner, H., and Brownell, H. H. (1986). *Right Hemisphere Communication Battery.* Boston, MA: Psychology Service.

Gibbs Jr, R. W., & Colston, H. L. (2006). Figurative Language. In *Handbook of Psycholinguistics* (pp. 835-861). Academic Press. https://doi.org/10.1016/B978-012369374-7/50022-5

Gibbs, R. W. (1999). Figurative Language. In R. A. Wilson (Ed.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 314–315). Cambridge MA: MIT Press.

Gibbs, R. W., Jr., & Beitel, D. (1995). What proverb understanding reveals about how people think. *Psychological Bulletin, 118,* 133–154. https://doi.org/10.1037/0033-2909.118.1.133

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM Mathematics Education, 47*(5), 771-782. https://doi.org/10.1007/s11858-014-0659-y

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8*, 183-206. https://doi.org/10.1515/cogl.1997.8.3.183

Giora, R., Gazal, O., Goldstein, I., Fein, O., & Stringaris, A. (2012). Salience and context: Interpretation of metaphorical and literal language by young adults diagnosed with Asperger's syndrome. *Metaphor and Symbol, 27*(1), 22-54. https://doi.org/10.1080/10926488.2012.638823

Giora, R., Gazal, O., Goldstein, I., Fein, O., & Stringaris, A. (2012). Salience and context: Interpretation of metaphorical and literal language by young adults diagnosed with Asperger's syndrome. *Metaphor and Symbol*, *27*(1), 22-54. https://doi.org/10.1080/10926488.2012.638823

Giora, R., Zaidel, E., Soroker, N., Batori, G., & Kasher, A. (2000). Differential effects of right- and left-hemisphere damage on understanding sarcasm and metaphor. *Metaphor and Symbol*, *15*(1-2), 63-83. https://doi.org/10.1080/10926488.2000.9678865

Gold, R., & Faust, M. (2010). Right hemisphere dysfunction and metaphor comprehension in young adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *40*(7), 800-811. https://doi.org/10.1007/s10803-009-0930-1

Golden, C. J., Kushner, T., Lee, B., & McMorrow, M. A. (1997). Searching for the meaning of the Category Test and the Wisconsin Card Sort Test: A comparative. *International Journal of Neuroscience, 93*, 141–150. https://doi.org/10.3109/00207459808986419

Gorham, D. R. (1956a). A proverbs test for clinical and experimental use. *Psychological Reports*. https://doi.org/10.2466/PR0.2.1-12

Gorham, D. R. (1956b). Use of the proverbs test for differentiating schizophrenics from normals. *Journal of Consulting Psychology*, *20*(6), 435-440. https://doi.org/10.1037/h0042949

Gottman, J., Gonso, J., & Rasmussen, B. (1975). Social interaction, social competence, and friendship in children. *Child Development*, 709-718. https://doi.org/10.2307/1128569

Graf, P., Uttl, B., & Tuokko, H. (1995). Color- and picture-word Stroop tests: Performance changes in old age. *Journal of Clinical and Experimental Neuropsychology, 17*, 390–415. https://doi.org/10.1080/01688639508405132

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of impairment and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology*, *39*, 404–411. https://doi.org/10.1037/h0059831

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*(3), 218. https://doi.org/10.1037/1040-3590.6.3.218

Greve, K. W., Bianchini, K. J., Hartley, S. M., & Adams, D. (1999). The Wisconsin Card Sorting Test in stroke rehabilitation: Factor structure and relationship to outcome. *Archives of Clinical Neuropsychology, 14*, 497–509. https://doi.org/10.1093/arclin/14.6.497

Greve, K. W., Ingram, F., & Bianchini, K. J. (1998). Latent structure of the Wisconsin Card Sorting Test in a clinical sample. *Archives of Clinical Neuropsychology, 13*, 597–609. https://doi.org/10.1093/arclin/13.7.597

Haas, M. H., Chance, S. A., Cram, D. F., Crow, T. J., Luc, A., & Hage, S. (2015). Evidence of pragmatic impairments in speech and proverb interpretation in schizophrenia. *Journal of Psycholinguistic Research*, *44*(4), 469-483. https://doi.org/10.1007/s10936-014-9298-2

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis: Pearson New International Edition.* Essex: Pearson Education Limited, *1*(2).

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications, Inc.

Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101-119. https://doi.org/10.1016/0010-0277(93)90026-R

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2016). *Acceptance and Commitment Therapy, Second Edition: The Process and Practice of Mindful Change (Second Edition).* The Guilford Press.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin Card Sorting Test (WCST) Manual, revised and expanded.* Odessa, FL: Psychological Assessment Resources.

Helland, W. A., Lundervold, A. J., Heimann, M., & Posserud, M. B. (2014). Stable associations between behavioral problems and language impairments across childhood–The importance of pragmatic language problems. *Research in Developmental Disabilities*, *35*(5), 943-951. https://doi.org/10.1016/j.ridd.2014.02.016

Hillert, D. G., & Buračas, G. T. (2009). The neural substrates of spoken idiom comprehension. *Language and Cognitive Processes*, *24*(9), 1370-1391. https://doi.org/10.1080/01690960903057006

Hyde, J. S. (2005) The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 53–69. https://doi.org/10.1037/0033-2909.104.1.53

Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155. https://doi.org/10.1037/0033-2909.107.2.139

Iakimova, G., Passerieux, C., & Hardy-Bayl, M. (2006). Interpretation of ambiguous idiomatic statements in schizophrenic and depressive patients. evidence for common and differential cognitive patterns. *Psychopathology, 39*(6), 277-285. https://doi.org/10.1159/000095732

Iakimova, G., Passerieux, C., & Hardy-Baylé, M. C. (2006). Interpretation of ambiguous idiomatic statements in schizophrenic and depressive patients. Evidence for common and differential cognitive patterns. *Psychopathology, 39*(6), 277-285. https://doi.org/10.1159/000095732

Iskandar, S. (2014). *The metaphor interpretation test: Cognitive processes involved and age group differences in performance.* Unpublished Ph.D. dissertation, University of Windsor

Iskandar, S., & Baird, A. D. (2014). The role of working memory and divided attention in metaphor interpretation. *Journal of Psycholinguistic Research,* 43, 555-568. https://doi.org/10.1007/s10936-013-9267-1

Ivnik, R. J., Malec, J. F., Smith, G. E., & Tangalos, E. G. (1996). Neuropsychological test norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, Stroop, TMT, and JLO. *The Clinical Neuropsychologist, 10*, 262–278. https://doi.org/10.1080/13854049608406689

Jackendoff, R. (1995). The boundaries of the lexicon. *Idioms: Structural and Psychological Perspectives*, *133*, 165.

Joanette, Y., & Brownell, H. (Eds.). (2012). *Discourse Ability and Brain Damage: Theoretical and Empirical Perspectives*. Springer Science & Business Media.

Joanette, Y., Goulet, P., Hannequin, D., & Boeglin, J. (1990). *Right Hemisphere and Verbal Communication*. Springer-Verlag Publishing. https://doi.org/10.1007/978-1-4612-4460-8

Johnson, A. T. (1996). Comprehension of metaphors and similes: A reaction time study. *Metaphor and Symbolic Activity, 11*(2), 145-159. https://doi.org/10.1207/s15327868ms1102_3

Jones, J., & Stone, C. A. (1989). Metaphor comprehension by language learning disabled and normally achieving adolescent boys. *Learning Disability Quarterly, 12*(4), 251- 60. https://doi.org/10.2307/1510208

Joormann, J. (2010). Cognitive inhibition and emotion regulation in depression. *Current Directions in Psychological Science,* 19(3), 161-166. https://doi.org/10.1177/0963721410370293

Kaiser, N. C., Lee, G. J., Lu, P. H., Mather, M. J., Shapira, J., Jimenez, E., & Mendez, M. F. (2013). What dementia reveals about proverb interpretation and its neuroanatomical correlates. *Neuropsychologia, 51*(9), 1726-1733. https://doi.org/10.1016/j.neuropsychologia.2013.05.021

Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbol*, *3*(4), 191-214. https://doi.org/10.1207/s15327868ms0304_1

Kaufman, A. S., & Lichtenberger, E. O. (2005). *Assessing Adolescent and Adult Intelligence.* John Wiley & Sons.

Kaya, F., Delen, E., & Bulut, O. (2012). Test review: Shipley-2 manual. *Journal of Psychoeducational Assessment, 30,* 593-597. https://doi.org/10.1177/0734282912440852

Kelly, W.A. (2007), *Cambridge Idioms Dictionary (Second Edition).* Cambridge University Press. https://doi.org/10.1108/09504120710744475

Kemple, K., Speranza, H., & Hazen, N. (1992). Cohesive discourse and peer acceptance: Longitudinal relations in the preschool years. *Merrill-Palmer Quarterly (1982),* 364-381.

Kempler, D., Van Lancker, D., & Read, S. (1988). Proverb and idiom comprehension in Alzheimer disease. *Alzheimer Disease and Associated Disorders*, *2*(1), 38-49. https://doi.org/10.1097/00002093-198802010-00005

Kempler, D., VanLancker, D., Marchman, V., & Bates, E. (1999). Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology*, *15*(3), 327-349. https://doi.org/10.1080/87565649909540753

Ketelaars, M. P., Cuperus, J. M., van Daal, J., Jansonius, K., & Verhoeven, L. (2009). Screening for pragmatic language impairment: The potential of the children's communication checklist. *Research in Developmental Disabilities*, *30*(5), 952-960. https://doi.org/10.1016/j.ridd.2009.01.006

Kiang, M., Light, G. A., Prugh, J., Coulson, S., Braff, D. L., & Kutas, M. (2007). Cognitive, neurophysiological, and functional correlates of proverb interpretation abnormalities in schizophrenia. *Journal of the International Neuropsychological Society*, *13*(4), 653-663. https://doi.org/10.1017/S1355617707070816

Kircher, T. T. J., Leube, D. T., Erb, M., Grodd, W., & Rapp, A. M. (2007). Neural correlates of metaphor processing in schizophrenia. *NeuroImage, 34*(1), 281-289. https://doi.org/10.1016/j.neuroimage.2006.08.044

Laiacona, M., Inzaghi, M. G., De Tanti, A., & Capitani, E. (2000). Wisconsin card sorting test: A new global score, with Italian norms and its relationship with the Weigl sorting test. *Neurological Science, 21*, 279–291. https://doi.org/10.1007/s100720070065

Lauro, L. J. R., Tettamanti, M., Cappa, S. F., & Papagno, C. (2008). Idiom comprehension: a prefrontal task?. *Cerebral Cortex*, *18*(1), 162-170. https://doi.org/10.1093/cercor/bhm042

Lee, R. F., & Kamhi, A. G. (1990). Metaphoric competence in children with learning disabilities. *Journal of Learning Disabilities*, *23*(8), 476-482. https://doi.org/10.1177/002221949002300805

Lee, S. S., & Dapretto, M. (2006). Metaphorical vs. literal word meanings: FMRI evidence against a selective role of the right hemisphere. *NeuroImage*, *29*(2), 536- 544. https://doi.org/10.1016/j.neuroimage.2005.08.003

Levin, B. E., Llabre, M. M., & Weiner, W. J. (1989). Cognitive impairments associated with early Parkinson's disease. *Neurology*, *39*(4), 557-557. https://doi.org/10.1212/WNL.39.4.557

Lewis, F. M., Woodyatt, G. C., & Murdoch, B. E. (2008). Linguistic and pragmatic language skills in adults with autism spectrum disorder: A pilot study. *Research in Autism Spectrum Disorders*, *2*(1), 176-187. https://doi.org/10.1016/j.rasd.2007.05.002

Leyhe, T., Saur, R., Eschweiler, G. W., & Milian, M. (2011). Impairment in proverb interpretation as an executive function deficit in patients with amnestic mild cognitive impairment and early Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders Extra*, *1*(1), 51-61. https://doi.org/10.1159/000323864

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory and Cognition, 36,* 1103–1121. https://doi.org/10.3758/MC.36.6.1103

Lodge, J. K. (2013). *The Concurrent Validity of the Shipley-2 and the WAIS-IV*. Unpublished Ph.D. dissertation, Wright State University.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 126. https://doi.org/10.1037/0278-7393.14.1.126

Marin, G., & Gamba, R. J. (1996). A new measurement of acculturation for Hispanics: The Bidimensional Acculturation Scale for Hispanics (BAS*). Hispanic Journal of Behavioral Sciences, 18*(3), 297-316. https://doi.org/10.1177/07399863960183002

Marini, Z., & Case, R. (1994). The development of abstract reasoning about the physical and social world. *Child Development*, *65*(1), 147-159. https://doi.org/10.2307/1131372

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language, 85*(3), 451-466. https://doi.org/10.1016/S0093-934X(03)00070-1

Martin, J. D., Blair, G. E., & Vickers, D. M. (1979). Correlation of the quick word test and wide range vocabulary test with the Shipley-Institute of Living Scale. *Educational and Psychological Measurement, 39*(4), 935-937. https://doi.org/10.1177/001316447903900430

Mashal, N., & Faust, M. (2010). The effects of metaphoricity and presentation style on brain activation during text comprehension. *Metaphor and Symbol, 25*(1), 19-33. https://doi.org/10.1080/10926480903538464

Mashal, N., Faust, M., & Hendler, T. (2005). The role of the right hemisphere in processing nonsalient metaphorical meanings: Application of principal components analysis to fMRI data. *Neuropsychologia, 43*(14), 2084-2100. https://doi.org/10.1016/j.neuropsychologia.2005.03.019

Mason, R. A., & Just, M. A. (2009). The role of the theory-of-mind cortical network in the comprehension of narratives. *Language and Linguistics Compass, 3*(1), 157-174. https://doi.org/10.1111/j.1749-818X.2008.00122.x

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition, and executive functions. *Language Learning and Development, 14*(3), 186-223. https://doi.org/10.1080/15475441.2018.1455584

May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 363–379. https://doi.org/10.1037/0096-1523.24.2.363

McCaffrey, R. J., Duff, K., & Westervelt, H. J. (2000). *Practitioner's Guide to Evaluating Change with Intellectual Assessment Instruments.* Springer. https://doi.org/10.1007/978-1-4615-4279-7

McDonald, C. R., Delis, D. C., Kramer, J. H., Tecoma, E. S., & Iragui, V. J. (2008). A componential analysis of proverb interpretation in patients with frontal lobe epilepsy and temporal lobe epilepsy: Relationships with disease-related factors. *The Clinical Neuropsychologist, 22*(3), 480-496. https://doi.org/10.1080/13854040701363828

McDonald, S. (1999). Exploring the process of inference generation in sarcasm: A review of normal and clinical studies. *Brain and Language, 68*(3), 486-506. https://doi.org/10.1006/brln.1999.2124

McDonald, S. (2008). Frontal lobes and language. In *Handbook of the Neuroscience of Language* (pp. 289-297). Elsevier. https://doi.org/10.1016/B978-008045352-1.00028-8

McDonald, S., & Pearce, S. (1996). Clinical insights into pragmatic theory: Frontal lobe deficits and sarcasm. *Brain and Language*, *53*(1), 81-104. https://doi.org/10.1006/brln.1996.0038

Meadow, A., Greenblatt, M., & Solomon, H. C. (1953). "Looseness of association" and impairment in abstraction in schizophrenia. *Journal of Nervous and Mental Disease*. https://doi.org/10.1097/00005053-195307000-00003

Medin, T., Ross, H., & Markman, A. B. (2002). *Cognitive Psychology (Third Edition).* New York: Wiley.

Mendoza, R. H. (1989). An empirical scale to measure type and degree of acculturation in Mexican American adolescents and adults. *Journal of Cross-Cultural Psychology, 20*(4), 372-385. https://doi.org/10.1177/0022022189204003

Merriam-Webster. (n.d.). *What is figurative language?* Merriam-Webster. Retrieved June 21, 2022, from https://www.merriam-webster.com/words-at-play/figurative-language

Mieder, W. (1978). The use of proverbs in psychological testing. *Journal of the Folklore Institute*, *15*(1), 45-55. https://doi.org/10.2307/3814137

Monetta, L., & Pell, M. D. (2007). Effects of verbal working memory deficits on metaphor comprehension in patients with Parkinson's disease. *Brain and Language, 101*(1), 80-89. https://doi.org/10.1016/j.bandl.2006.06.007

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 520–538. https://doi.org/10.1037/0096-1523.17.2.520

Muran, J. C., & DiGiuseppe, R. A. (1990). Towards a cognitive formulation of metaphor use in psychotherapy. *Clinical Psychology Review, 10*(1), 69-85. https://doi.org/10.1016/0272-7358(90)90107-L

Murphy, P., Shallice, T., Robinson, G., MacPherson, S. E., Turner, M., Woollett, K., & Cipolotti, L. (2013). Impairments in proverb interpretation following focal frontal lobe lesions. *Neuropsychologia*, *51*(11), 2075-2086. https://doi.org/10.1016/j.neuropsychologia.2013.06.029

Murphy, S. M., Faulkner, D. M., & Farley, L. R. (2014). The behaviour of young children with social communication disorders during dyadic interaction with peers. *Journal of Abnormal Child Psychology, 42*(2), 277-289. https://doi.org/10.1007/s10802-013-9772-6

Nakagawa, Y., Tanabe, H., Ikeda, M., Kazui, H., Ito, K., Inoue, N., & Shiraishi, J. (1993). Completion phenomenon in transcortical sensory aphasia. *Behavioural Neurology, 6*(3), 135-142. https://doi.org/10.1155/1993/267103

Nenonen, M., Niemi, J., & Laine, M. (2002). Representation and processing of idioms: Evidence from aphasia. *Journal of Neurolinguistics, 15*(1), 43-58. https://doi.org/10.1016/S0911-6044(01)00005-7

Nippold, M. A., & Duthie, J. K. (2003). Mental imagery and idiom comprehension: A comparison of school-age children and adults. *Journal of Speech, Language, and Hearing Research, 46*(4), 788-799. https://doi.org/10.1044/1092-4388(2003/062)

Nippold, M. A., & Fey, S. H. (1983). Metaphoric understanding in preadolescents having a history of language acquisition difficulties. *Language, Speech, and Hearing Services in Schools, 14*(3), 171-180. https://doi.org/10.1044/0161-1461.1403.171

Nippold, M. A., Uhden, L. D., & Schwarz, I. E. (1997). Proverb explanation through the lifespan: A developmental study of adolescents and adults. *Journal of Speech, Language & Hearing Research, 40*(2), 245–253. https://doi.org/10.1044/jslhr.4002.245

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, *70*(3), 491-538. https://doi.org/10.1353/lan.1994.0007

Nunnally, J. C. (1978). *Psychometric Theory: Second Edition.* New York: McGraw-Hill.

Oliveri, M., Romero, L., & Papagno, C. (2004). Left but not right temporal involvement in opaque idiom comprehension: A repetitive transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, *16*(5), 848-855. https://doi.org/10.1162/089892904970717

Ortony, A. (1993). *Metaphor and Thought (Second Edition).* Cambridge, England; New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781139173865

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior, 17*(4), 465-477. https://doi.org/10.1016/S0022-5371(78)90283-9

Paltridge, B. (2021). *Discourse Analysis: An Introduction*. Bloomsbury Publishing.

Paolo, A. M., Axelrod, B. N., & Troster, A. I. (1996). Test-retest stability of the Wisconsin Card Sorting Test. *Assessment, 3*, 137–143. https://doi.org/10.1177/107319119600300205

Papagno, C. (2001). Comprehension of metaphors and idioms in patients with Alzheimer's Disease. *Brain, 124*(7), 1450-1460. https://doi.org/10.1093/brain/124.7.1450

Papagno, C. (2010). Idiomatic language comprehension: Neuropsychological evidence. In *Neuropsychology of Communication* (pp. 111-129). Springer, Milano. https://doi.org/10.1007/978-88-470-1584-5_6

Papagno, C., & Caporali, A. (2007). Testing idiom comprehension in aphasic patients: The effects of task and idiom type. *Brain and Language*, *100*(2), 208-220.

Papagno, C., & Genoni, A. (2004). The role of syntactic competence in idiom comprehension: a study on aphasic patients. *Journal of Neurolinguistics*, *17*(5), 371-382. https://doi.org/10.1016/j.jneuroling.2003.11.002

Papagno, C., & Vallar, G. (2001). Understanding metaphors and idioms: A single-case neuropsychological study in a person with down syndrome. *Journal of the International Neuropsychological Society, 7*(4), 516-527. https://doi.org/10.1017/S1355617701744086

Papagno, C., Cappa, S. F., Forelli, A., Garavaglia, G., Laiacona, M., Capitani, E., & Vallar, G. (1995). Non-literal language comprehension: standardization of a metaphor and idiom comprehension test. *Archivio di Psicologia, Neurologia e Psichiatria*, *56*(4), 402-420.

Papagno, C., Curti, R., Rizzo, S., Crippa, F., & Colombo, M. R. (2006). Is the right hemisphere involved in idiom comprehension? A neuropsychological study. *Neuropsychology*, *20*(5), 598. https://doi.org/10.1037/0894-4105.20.5.598

Papagno, C., Lucchelli, F., Muggia, S., & Rizzo, S. (2003). Idiom comprehension in Alzheimer's disease: The role of the central executive. *Brain*, *126*(11), 2419-2430. https://doi.org/10.1093/brain/awg243

Pendleton, M. G., & Heaton, R. K. (1982). A comparison of the Wisconsin Card Sorting Test and the Category Test. *Journal of Clinical Psychology, 38*, 392–396. https://doi.org/10.1002/1097-4679(198204)38:2<392:AID-JCLP2270380232>3.0.CO;2-W

Phelps-Terasaki, D., & Phelps-Gunn, T. (2007). *Test of Pragmatic Language–Second Edition*. East Moline: Linguisystems.

Pineda, D. A., & Merchan, V. (2003). Executive function in young Colombian adults. *International Journal of Neuroscience, 113,* 397–410. https://doi.org/10.1080/00207450390162164

Pobric, G., Mashal, N., Faust, M., & Lavidor, M. (2008). The role of the right cerebral hemisphere in processing novel metaphoric expressions: A transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience, 20*(1), 170-181. https://doi.org/10.1162/jocn.2008.20005

Prutting, C. A., and Kirchner, D. M. (1987). A clinical appraisal of the pragmatic aspects of language. *Journal of Speech and Hearing Disorders, 52*, 105–119. https://doi.org/10.1044/jshd.5202.105

Psychological Corporation. (1997). *Wechsler Adult Intelligence Scale–Third Edition, and Wechsler Memory Scale–Third Edition Technical Manual*. San Antonio, TX: The Psychological Corporation.

Pukrop, R., Matuschek, E., Ruhrmann, S., Brockhaus-Dumke, A., Tendolkar, I., Bertsch, A., & Klosterkotter, J. (2003). Dimensions of working memory dysfunction in schizophrenia. *Schizophrenia Research, 62*, 259–268. https://doi.org/10.1016/S0920-9964(02)00427-9

Pynte, J., Besson, M., Robichon, F., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language, 55*(3), 293- 316. https://doi.org/10.1006/brln.1996.0107

Rapp, A. M., & Wild, B. (2011). Nonliteral language in Alzheimer dementia: A review. *Journal of the International Neuropsychological Society, 17*(2), 207-218. https://doi.org/10.1017/S1355617710001682

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. J. (2004). Neural correlates of metaphor processing. *Cognitive Brain Research, 20*(3), 395-402. https://doi.org/10.1016/j.cogbrainres.2004.03.017

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. J. (2007). Laterality in metaphor processing: Lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain and Language, 100*(2), 142-149. https://doi.org/10.1016/j.bandl.2006.04.004

Rassiga, C., Lucchelli, F., Crippa, F., & Papagno, C. (2009). Ambiguous idiom comprehension in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 31*(4), 402-411. https://doi.org/10.1080/13803390802220019

Regard, M. (1981). *Cognitive Rigidity and Flexibility: A Neuropsychological Study.* Unpublished Ph.D. dissertation, University of Victoria.

Rehmel, J. L., Brown, W. S., & Paul, L. K. (2016). Proverb comprehension in individuals with agenesis of the corpus callosum. *Brain and Language, 160*, 21-29. https://doi.org/10.1016/j.bandl.2016.07.001

Rinaldi, M. C., Marangolo, P., & Baldassarri, F. (2004). Metaphor comprehension in right brain-damaged patients with visuo-verbal and verbal material: A dissociation (re) considered. *Cortex, 40*(3), 479-490. https://doi.org/10.1016/S0010-9452(08)70141-2

Rizzo, S., Sandrini, M., & Papagno, C. (2007). The dorsolateral prefrontal cortex in idiom interpretation: An rTMS study. *Brain Research Bulletin, 71*(5), 523-528. https://doi.org/10.1016/j.brainresbull.2006.11.006

Root, William T. (1921). A socio-psychological study of fifty-three supernormal children. *Psychological Monographs, 29*(4), 85. https://doi.org/10.1037/h0093136

Rubin, M. (2017). Do p values lose their meaning in exploratory analyses? It depends how you define the familywise error rate. *Review of General Psychology,* 21(3), 269-275. https://doi.org/10.1037/gpr0000123

Saklofske, D. H., Chelune, G. J., Heaton, R. K., Ivnik, R. J., Bornstein, R., Prifitera, A., & Ledbetter, M. F. (2003). *Clinical Interpretation of the WAIS-III and WMS-III.* Academic Press.

Schettino, A., Lauro, L. R., Crippa, F., Anselmetti, S., Cavallaro, R., & Papagno, C. (2010). The comprehension of idiomatic expressions in schizophrenic patients. *Neuropsychologia*, *48*(4), 1032-1040. https://doi.org/10.1016/j.neuropsychologia.2009.11.030

Schmidt, G. L., & Seger, C. A. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity, and difficulty. *Brain and Cognition, 71*(3), 375- 386. https://doi.org/10.1016/j.bandc.2009.06.001

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment,* 19(1), 21-30. https://doi.org/10.1177/1073191111428764

Schweizer, K., & DiStefano, C. (Eds.). (2016). *Principles and methods of test construction: Standards and recent advances.* Hogrefe. https://doi.org/10.1027/00449-000

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry,* 405-414. https://doi.org/10.2307/2785779

Searle, J. R. (1979). What is an intentional state? *Mind*, *88*(349), 74-92. https://doi.org/10.1093/mind/LXXXVIII.1.74

Sherman, E. M., Strauss, E., Spellacy, F., & Hunter, M. (1995). Construct validity of WAIS-R factors: Neuropsychological correlates in adults referred for possible head injury. *Psychological Assessment, 7*, 440–444. https://doi.org/10.1037/1040-3590.7.4.440

Shibata, M., Abe, J. i., Terao, A., & Miyamoto, T. (2007). Neural mechanisms involved in the comprehension of metaphoric and literal sentences: An fMRI study. *Brain Research, 1166*, 92-102. https://doi.org/10.1016/j.brainres.2007.06.040

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9*, 371-377. https://doi.org/10.1080/00223980.1940.9917704

Shipley, W. C. (1946). *Institute of Living Scale.* Los Angeles, CA: Western Psychological Services.

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2*. Los Angeles, CA: Western Psychological Services. https://doi.org/10.1037/t48948-000

Sotillo, M., Carretié, L., Hinojosa, J. A., Tapia, M., Mercado, F., López-Martín, S., & Albert, J. (2004). Neural activity associated with metaphor comprehension: Spatial analysis. *Neuroscience Letters, 373*(1), 5-9. https://doi.org/10.1016/j.neulet.2004.09.071

Speake, J. (Ed.). (2015). *Oxford Dictionary of Proverbs*. OUP Oxford. https://doi.org/10.1093/acref/9780198734901.001.0001

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60*(9), 950–958. https://doi.org/10.1037/0003-066X.60.9.950

Sponheim, S. R., Surerus-Johnson, C., Leskela, J., & Dieperink, M. E. (2003). Proverb interpretation in schizophrenia: the significance of symptomatology and cognitive processes. *Schizophrenia Research, 65*(2-3), 117-123. https://doi.org/10.1016/S0920-9964(02)00525-X

St Clair, M. C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional, and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders, 44*(2), 186-199. https://doi.org/10.1016/j.jcomdis.2010.09.004

Stachowiak, F. J., Huber, W., Poeck, K., & Kerschensteiner, M. (1977). Text comprehension in aphasia. *Brain and Language, 4*(2), 177-195. https://doi.org/10.1016/0093-934X(77)90016-5

Stemmer B. (2008). Neuropragmatics: Disorders and Neural Systems, in *Handbook of the Neuroscience of Language*, (Eds.) Stemmer B., Whitaker H. A. (New York, NY: Elsevier), 175–187. https://doi.org/10.1016/B978-0-08-045352-1.00017-3

Stemmer, B., & Whitaker, H. A. (Eds.). (2008). *Handbook of the Neuroscience of Language*. Academic Press.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (Third Edition).* Oxford University Press.

Stringaris, A. K., Medford, N. C., Giampietro, V., Brammer, M. J., & David, A. S. (2007). Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. *Brain and Language, 100(2),* 150-162. https://doi.org/10.1016/j.bandl.2005.08.001

Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology, 18,* 643–662. https://doi.org/10.1037/h0054651

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research, 63*(34), 289- 298. https://doi.org/10.1007/s004269900007

Stuss, D. T., Gallup Jr, G. G., & Alexander, M. P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain*, *124*(2), 279-286. https://doi.org/10.1093/brain/124.2.279

Thoma, P., Hennecke, M., Mandok, T., Wähner, A., Brüne, M., Juckel, G., & Daum, I. (2009). Proverb comprehension impairments in schizophrenia are related to executive dysfunction. *Psychiatry Research*, *170*(2-3), 132-139. https://doi.org/10.1016/j.psychres.2009.01.026

Thorsen, C., Gustafsson, J. E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology*, *84*(4), 556-570. https://doi.org/10.1111/bjep.12041

Titone, D. A., & Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbol*, *9*(4), 247-270. https://doi.org/10.1207/s15327868ms0904_1

Titone, D., Holzman, P. S., & Levy, D. L. (2002). Idiom processing in schizophrenia: literal implausibility saves the day for idiom priming. *Journal of Abnormal Psychology*, *111*(2), 313. https://doi.org/10.1037/0021-843X.111.2.313

Uekermann, J., Thoma, P., & Daum, I. (2008). Proverb interpretation changes in aging. *Brain and Cognition*, *67*(1), 51-57. https://doi.org/10.1016/j.bandc.2007.11.003

Utsumi, A. (2011). Computational exploration of metaphor comprehension processes using a semantic space model. *Cognitive Science*, *35*(2), 251–296. https://doi.org/10.1111/j.1551-6709.2010.01144.x

Van Lancker, D. R., & Kempler, D. (1987). Comprehension of familiar phrases by left-but not by right-hemisphere damaged patients. *Brain and language*, *32*(2), 265-277. https://doi.org/10.1016/0093-934X(87)90128-3

Vosniadou, S. (1987). Children and metaphors. *Child Development, 58*(3), 870-885. https://doi.org/10.2307/1130223

Waltz, D. L., & Pollack, J. B. (1985). Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science*, *9*(1), 51–74. https://doi.org/10.1016/S0364-0213(85)80009-4

Wang, J. J., & Kaufman, A. S. (1993). Changes in fluid and crystallized intelligence across the 20-to 90-year age range on the K-BIT. *Journal of Psychoeducational Assessment*, 11(1), 29-37. https://doi.org/10.1177/073428299301100104

Watts, A. (1950). *The Language of Mental Development.* London: Harrap.

Wechsler, D. (1958). The measurement and appraisal of adult intelligence. https://doi.org/10.1037/11167-000

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition/Wechsler Memory Scale—Third Edition Technical Manual.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition.* New York: Psychological Corporation. https://doi.org/10.1037/t15169-000

Weiss, J. L., & Schell, R. E. (1991). Estimating WAIS-R IQ from the Shipley institute of living scale: A replication. *Journal of Clinical Psychology*, *47*(4), 558-562. https://doi.org/10.1002/1097-4679(199107)47:4<558:AID-JCLP2270470414>3.0.CO;2-W

Whitehouse, A. J., Watt, H. J., Line, E. A., & Bishop, D. V. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language & Communication Disorders*, *44*(4), 511-528. https://doi.org/10.1080/13682820802708098

Wiig, E., & Secord, W. A. (2014*). Clinical Evaluation of Language Fundamentals–Fifth Edition Metalinguistics (CELF-5 Metalinguistics).* Pearson Psych Corp.

Winkler, I. (1980). *Incremental Validity of Short IQ Tests*. Unpublished Ph.D. dissertation, The University of Iowa.

Winner, E., & Gardner, H. (1977). The comprehension of metaphor in brain-damaged patients. *Brain, 100*(4), 717-729. https://doi.org/10.1093/brain/100.4.717

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language, 62*(1), 89-106. https://doi.org/10.1006/brln.1997.1889

Wolstencroft, J., Robinson, L., Srinivasan, R., Kerry, E., Mandy, W., & Skuse, D. (2018). A systematic review of group social skills interventions, and meta-analysis of outcomes, for children with high functioning ASD. *Journal of Autism and Developmental Disorders*, *48*(7), 2293-2307. https://doi.org/10.1007/s10803-018-3485-1

Yerys, B. E., Antezana, L., Weinblatt, R., Jankowski, K. F., Strang, J., Vaidya, C. J., Schultz R. T., Gaillard, W. D., & Kenworthy, L. (2015). Neural correlates of set-shifting in children with autism. *Autism Research,* 8(4), 386-397. https://doi.org/10.1002/aur.1454

Yi, Y. G., Kim, D. Y., Shim, W. H., Oh, J. Y., Kim, S. H., & Kim, H. S. (2017). Neural correlates of Korean proverb processing: A functional magnetic resonance imaging study. *Brain and Behavior*, *7*(10), e00829. https://doi.org/10.1002/brb3.829

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94. https://doi.org/10.20982/tqmp.09.2.p079

Zempleni, M. Z., Haverkort, M., Renken, R., & Stowe, L. A. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *NeuroImage*, *34*(3), 1280-1291. https://doi.org/10.1016/j.neuroimage.2006.09.049

APPENDICES

APPENDIX A: FIGURATIVE LANGUAGE INTERPRETATION TASK PILOT ITEMS WITH DEFINITIONS

The items chosen for this measure were selected from previous studies and publications on figurative language (Benjafield et al., 1993; Billow, 1975; Katz et al., 1988; Campbell & Raney, 2016; Titone & Connine, 1994; Libben & Titone, 2008; Tanner & Bulkes, 2017; Aussie Slang Dictionary, 2022). The example sentences (when required) and definitions used were based on previous publications and available dictionaries (Speake, 2015; Iskandar & Baird, 2013; Iskandar, 2014; Ayto & Oxford University Press, 2020; Kelly, 2007; Aussie Slang Dictionary, 2022).

Familiar (American English) Idioms

1. I thought she really liked me, but the next day she *gave me the cold shoulder*.
    a. To unintentionally ignore someone or treat someone in an unfriendly way.

2. I was going to try bungee jumping, but I *got cold feet.*
    a. To feel too frightened to do something that you had planned to do.

3. Emails that criticize others can land you *in hot water*, so be careful what you write.
    a. in a difficult situation in which you are likely to be punished.

4. Stop *pulling my leg* – you didn't have lunch with JLo!
    a. To tell someone something that is not true as a way of joking with the person.

5. They only met three weeks ago – isn't it *jumping the gun* to start talking about marriage?
    a. To act too soon or before the right time.

6. You're *playing with fire* if you try to cheat on the test.
    a. To do something that could cause you trouble later.

7. Environmentalists have no political *axe to grind* – they just want to save the planet.
    a. To have a strong personal opinion about something that you want people to accept and that is the reason why you do something.

8. Many consumers are still *on the fence*, waiting for a less expensive electric vehicle to come along.
    a. Not able to decide something.

9. Don't *waste your breath* arguing with him.
    a. To say something that will likely be ignored.

10. Someone suggested that we play a party game to *break the ice.*
    a. To do or say something that makes people who do not know each other feel more comfortable.

11. If she gets caught lying to me, then she'll just have to *face the music.*
    a. Be confronted with the unpleasant consequences of your actions.

12. We said he left "to pursue other interests" to let him *save face,* but actually we fired him.
    a. Able to avoid humiliation.

13. We sat out on the porch to spend time together, just *shooting the breeze.*
    a. To spend time talking about things that are not important.

14. I'm feeling a little *under the weather*; I don't think I will go out today.
    a. To not feel well.

15. She passed the exam with *flying colors.*
    a. To complete something very easily.

Unfamiliar (Australian English) Idioms

16. She has some great ideas, but by God can she *carry on like a pork chop* about nothing.
    a. To behave foolishly, to make a fuss, to complain, or to rant.

17. Hey man, it's only a scratch, don't *get off your bike about it.*
    a. It is not the biggest problem in the world.

18. *Get off the grass* – you weren't on the news last night.
    a. Untrue or unbelievable information.

19. By the time customer service called me back I was totally *jack of it.*
    a. To have enough, to be over it.

20. Did you see that guy doing a wheelie at that intersection? Must have a few *kangaroos loose in the top!*
    a. To act, think, or behave in an eccentric, foolish, or nonsensical manner.

21. *Keep your shirt on,* he wasn't trying to make you mad.
    a. Hold your temper, stay calm.

22. My wife went off like *a frog in a sock* when she saw the mess I left in the garage.
    a. To lose one's temper.

23. I don't know what he's talking about, he's *lost the plot.*
    a. To be beyond normal reasoning.

24. You've got to *pull your socks up*; this just isn't good enough!
    a. To improve or do better.

25. You need to *pull your head in,* if they wanted you to know they'd tell you!

a. Worry about yourself, don't pry into other people's business/personal lives.

26. It's a bit of a mess right now, but it'll be *right as rain* come tomorrow.
    a. Everything will be OK.

27. I tried to make him understand, but that guy is as *thick as a brick.*
    a. To be unintelligent.

28. If you are going to spend this much time on the project, at least go and *make a bird of it.*
    a. To do something well or properly.

29. I think I'm going to need a new car soon, this one's *on the blink.*
    a. About to break-down, will not be useful for much longer.

30. Be careful, he's *lower than a snake's belly.*
    a. Someone that is not of good character.

Proverbs

31. *Do as I say, not as I do.*
    a. Follow my orders or instructions rather than my example.

32. *Love is blind.*
    a. Love can be random or illogical.

33. *You cannot have your cake and eat it too.*
    a. If you consume or spend something, you cannot still possess it.

34. *Nothing so certain as death.*
    a. No one evades the reach of death.

35. *As a man lives, so shall he die, as a tree falls, so shall it lie.*
    a. One must not change long-established beliefs in the face of death.

36. *Many strokes fell great oaks.*
    a. Limited strength, when persistently applied, can accomplish great feats.

37. *No pain, no gain.*
    a. Nothing is achieved without effort.

38. *Two heads are better than one.*
    a. The cooperation of another person in solving a problem or making a plan is a good idea.

39. *A bird in the hand is worth two in the bush.*

a. It is better to accept or be content with what one has rather than try to get more and risk losing everything.

40. *Every bread has its crust.*
    a. Even good things are not perfect.

41. *The grass is always greener on the other side of the fence.*
    a. Other people's lives or situations always seem better than your own.

42. *Too many cooks spoil the broth.*
    a. Having too many people involved in running an enterprise is damaging to its chances of success.

43. *Every cloud has a silver lining.*
    a. Every difficult or sad situation has a comforting or more hopeful aspect, even though this may not be immediately apparent.

44. *A watched pot never boils.*
    a. Watching and waiting impatiently for something to occur usually seems to postpone the event.

45. *Don't cry over spilt milk.*
    a. It is pointless to express regret or grief when it is too late to remedy the misfortune.

Metaphors

46. *The mind is a sponge.*
    a. The mind absorbs/soaks up/takes in/sucks in information/knowledge/facts.

47. *Faithful love is a tree standing through the stormiest hour.*
    a. Faithful love withstands/overcomes/work through/make it through/remains/endures difficult situations.

48. *The stars are signposts.*
    a. Functions as a guide; Provide direction/guidance; Point/lead/show the way; Navigate.

49. *Snow is winter's robes*.
    a. Snow covers/blankets/coats/dresses the ground/earth.

50. *Love is a flower.*
    a. Love blooms/grows/blossoms.

51. *Genes are blueprints.*
    a. Genes dictate/determine/lay out/map out how something/the body/a person will turn out.

52. *Hard work is a ladder.*
    a. Hard work helps you improve/lifts up/moves up; Can get you to higher levels/places/to the top.

53. *Education is a lantern.*
    a. Education sheds light/brightens/illuminates/enlightens/shines; Lights up the way/path/future/life/mind.

54. *A judge is a balance.*
    a. A judge considers/weighs/measures/mediates/determines/sees/evaluates all/both/two views/sides/the evidence before coming to a decision.

55. *Alcohol is a crutch.*
    a. Alcohol is something you lean on/depend on/rely on.

56. *A butterfly is a winged rainbow.*
    a. Colorful.

57. *A tree is an umbrella.*
    a. An umbrella shields/guards/covers/protects from the rain/sun; Provides shade/shelter/protection.

58. *The stars are fireflies.*
    a. Fireflies light up/blink/flicker/twinkle/sparkle/give off light at night.

59. *A pimple is the skin's volcano*.
    a. A pimple erupts/explodes/bursts/ruptures/blasts/emits.

60. *An autumn's storm is the funeral song of a dying year.*
    a. An autumn's storm signals/indicates/marks/represents/reflects/symbolizes the end of the year; Shows that the year is coming to an end/over.

APPENDIX B: FIGURATIVE LANGUAGE INTERPRETATION TASK SCORING GUIDELINES

Free Inquiry Responses

a. Verbal responses can vary according to the *Degree of Accuracy* and the *Degree of Abstraction*. These two dimensions of a response are scored independently or each other and the two scores are combined to form a single achievement score for each response. Consequently, free inquiry responses fall into five general types:

Description	Accuracy Score (max: 2)	Abstraction Score (max: 2)	Total Score (max: 4)
Accurate & Abstract	2	2	4
Partial & Abstract	1	2	3
Accurate & Concrete	2	0	2
Partial & Concrete	1	0	1
Incorrect/Concrete	0	0	0

b. Guidelines for rating *Degree of Accuracy*
    i. Accuracy is rated according to how well the meaning of the response "fits" the meaning of the statement, regardless of the level of abstraction of the response.
    ii. Accuracy is rated according to a 3-point scale:
        1. Mostly Accurate – 2 points
        2. Partially Accurate – 1 point
        3. Mostly Inaccurate – 0 points
    iii. Mostly Accurate: if all or most of the key elements of a statement are accurately interpreted or implied by the examinee's response—either abstractly or concretely—then 2 points are awarded for accuracy.
    iv. Partially Accurate: if some key elements of a statement are interpreted accurately, but other important elements are omitted or incorrectly interpreted, then 1 point is awarded for accuracy.
    v. Mostly Inaccurate: if none or few of the key elements are interpreted by the examinee's response, then 0 points are awarded for accuracy. Inaccurate responses (a) miss both the concrete or abstract meaning of the key elements in the statement or (b) use almost exclusively the same words as the statement.

c. Guidelines for Rating *Degree of Abstraction*
    i. Abstraction is rated according to the level of generalization represented by the response *without regard to its accuracy or inaccuracy.* The examiner must determine whether the response applies to more situations, people, or concepts than are reflected in the literal meaning of the original statement.
    ii. Abstraction is rated according to an either-or, 2-point scale:
        1. Mostly Abstract – 2 points
        2. Mostly Concrete – 0 points
    iii. Mostly Abstract: if the examinee attempts to provide an interpretation that applies to more situations, people, or concepts than does the literal meaning of the original statement, then 2 points are awarded for abstraction.

iv. Mostly Concrete: if the examinee's interpretation of a statement is concrete or literal, then 0 points are awarded for abstraction.

d. It is important to note that if an examinee's response scores a 0 for *Degree of Accuracy,* then their total score is also a 0, regardless of their *Degree of Abstraction.*

APPENDIX C: REVISED FIGURATIVE LANGUAGE INTERPRETATION TASK ITEMS WITH EXAMPLE RESPONSES FOR EACH POINT VALUE

The items chosen for this measure were selected from previous studies and publications on figurative language (Benjafield et al., 1993; Billow, 1975; Katz et al., 1988; Campbell & Raney, 2016; Titone & Connine, 1994; Libben & Titone, 2008; Tanner & Bulkes, 2017). The example sentences (when required) and definitions used were based on previous publications and available dictionaries (Speake, 2015; Iskandar & Baird, 2013; Iskandar, 2014; Ayto & Oxford University Press, 2020; Kelly, 2007), and were modified and expanded based on the results from Study One.

Idioms

1. Emails that criticize others can land you *in hot water,* so be careful what you write.
    - 4 Point: In a difficult situation in which you are likely to be punished; in trouble.
    - 3 Point: Bad situation; in danger; dangerous.
    - 2 Point: Gives a literal response about water and water temperature.

2. You're *playing with fire* if you try to cheat on the test.
    - 4 Point: To do something that could cause you trouble later; being risky; taking a risk.
    - 3 Point: Messing with something that you're not supposed to; dangerous.
    - 2 Point: Gives a literal response about fire and/or the dangers of fire.

3. Many consumers are still *on the fence*, waiting for a less expensive electric vehicle to come along.
    - 4 Point: Not able to decide something; debating; considering; deciding.
    - 3 Point: Thinking about it; waiting for more information; half-in; unsure.
    - 2 Point: Gives a literal response about a fence.

4. We sat out on the porch to spend time together, just *shooting the breeze.*
    - 4 Point: To spend time talking about things that are not important; small talk.
    - 3 Point: Hanging out; chilling; talking; relaxing.
    - 2 Point: Gives a literal interpretation about shooting or the weather conditions.

5. I'm feeling a little *under the weather*; I don't think I will go out today.
    - 4 Point: To not feel well; be sick/ill.
    - 3 Point: Not good; bad.
    - 2 Point: Gives a literal interpretation about the weather.

6. She passed the exam with *flying colors.*
    - 4 Point: To complete something very easily; very well; excellent (or synonym).
    - 3 Point: Perfect score; average/high score.
    - 2 Point: Gives a literal interpretation about colors or flying.

Proverbs

7. *Do as I say, not as I do.*
    - 4 Point: Follow/listen to my orders/commands/directions/instructions rather than my example/what I do.
    - 3 Point: Follow orders/commands/directions/rules; don't follow my actions; they don't follow their own rules; listen to what I say.
    - 2 Point: Repeats the phrase with minimal re-wording.

8. *Love is blind.*
    - 4 Point: Love can be random or illogical; makes you unable to see problems; ignore the flaws/concerns/red flags.
    - 3 Point: You do things you wouldn't normally do/act differently/do dumb things; you don't see it all/clearly.
    - 2 Point: You don't care what they look like; love isn't about looks; love is about personality.

9. *Nothing so certain as death.*
    - 4 Point: No one evades the reach of death; death is certain/will happen/going to happen; everyone dies.
    - 3 Point: Nothing is guaranteed/certain in life; anything can happen/is possible; whatever is going on is inevitable/guaranteed, just like death.
    - 2 Point: Describes death or dying.

10. *Many strokes fell great oaks.*
    - 4 Point: Limited strength, when persistently applied, can accomplish great feats; lots of effort accomplishes big tasks; great things require effort.
    - 3 Point: Repeatedly doing something will eventually lead to results; keep trying and you will succeed; big things require repeated attempts; you can chip away at things to get them done; it requires a lot of effort.
    - 2 Point: A tree will fall if you repeatedly chop at it; you can cut down trees.

11. *Every bread has its crust.*
    - 4 Point: Even good things are not perfect; everything has flaws/negative aspects.
    - 3 Point: There's always something you don't like/something bad; there are shortcomings/problems.
    - 2 Point: There are rough/tough outsides; there is an outer layer.

12. *The grass is always greener on the other side of the fence***.**
    - 4 Point: Other people's lives or situations always seem better/more appealing than your own; things seem better when you can't have them/they are out of reach; you don't appreciate what you have.
    - 3 Point: You want what you can't have; you are never satisfied; other people have better/nice things/situations.

- 2 Point: Things look better from afar; things/it look better on the other side.

13. *Every cloud has a silver lining.*
    - 4 Point: Every difficult or sad situation has a comforting or more hopeful aspect, even though this may not be immediately apparent; there is always a bright side; there are positive elements to every bad situation.
    - 3 Point: There's always good; There's good and bad to everything; try to be optimistic.
    - 2 Point: Clouds form different shapes/colors; clouds come with storms and bring rainbows.

14. *A watched pot never boils.*
    - 4 Point: Watching and waiting impatiently for something to occur usually seems to postpone the event; waiting for something to happen makes it feel like it's taking longer.
    - 3 Point: Be patient; if you worry, it won't happen/less likely to happen; anxiety makes time seem to drag on; sit back and let things take their course.
    - 2 Point: You have to look away for something to happen.

15. *Don't cry over spilt milk.*
    - 4 Point: It is pointless to express regret or grief when it is too late to remedy the misfortune; don't sweat the small stuff; don't get upset over small/little/trivial things.
    - 3 Point: Don't be dramatic; don't make it a big deal; don't get upset; you can't change things that are over.
    - 2 Point: Don't cry, just clean it up.

Metaphors

16. *Faithful love is a tree standing through the stormiest hour.*
    a. 4 Point: Faithful love withstands/overcomes/works through/makes it through difficult situations; Staying true despite difficult times.
    b. 3 Point: Love is hopeful/positive in bad moments; love gets you through/trumps all; someone stands by you when it's hard; love gets through things/anything.
    c. 2 Point: Has roots; deep roots; provides shelter; stands tall.

17. *The stars are signposts.*
    - 4 Point: Functions as a guide; Provides direction/guidance; Points/leads/shows the way; Navigate/Navigation.
    - 3 Point: Follow the directions; like a map; you can get a sense of direction.
    - 2 Point: You see stars in the sky; they can look like signs in the sky.

18. *Love is a flower.*
    - 4 Point: Love blooms/grows/blossoms.
    - 3 Point: Requires effort, nourishment, attention.
    - 2 Point: Delicate; beautiful; pretty; any physical description of a flower.

19. *Genes are blueprints.*
    - 4 Point: Genes dictate/determine/lay out/map out how you will be/who you are/what you are like.
    - 3 Point: Gives insight into how it's built/how to build you; make up who you are; make what you are born like; indicates certain aspects of you.
    - 2 Point: Repeats the phrase with minimal re-wording; they make you human; basic building instructions; make up your body.

20. *Education is a lantern.*
    - 4 Point: Education sheds light/brightens/illuminates/enlightens; Lights the way/path/future.
    - 3 Point: Provides light; the light of our world/future; helps you see the way/see things differently.
    - 2 Point: Education helps you learn; you learn a lot from education; helps you see in the dark.

21. *A judge is a balance.*
    - 4 Point: A judge considers/weighs/determines/evaluates all/both/two views/sides/evidence.
    - 3 Point: Weighs actions; weighs/attempts to balance good and bad/good and evil.
    - 2 Point: A judge is unbiased; provides justice; passes judgements; is a mediator; describes only a balance.

22. *Alcohol is a crutch.*
    - 4 Point: Alcohol is something you lean on/depend on/rely on; Enables people; Short-term/temporary relief/benefits/solution.
    - 3 Point: Used to get you through hard times; used to cope; helps you ignore things you should face; doesn't actually help the issue; used to medicate problems.
    - 2 Point: Is bad/not useful/harmful.